Emotions: the art of communication applied to virtual actors

Tanguy, Emmanuel Adrien Raymond

Award date:
2006

Awarding institution:
University of Bath

Emotions: the Art of Communication Applied to Virtual Actors

Emmanuel Adrien Raymond Tanguy

Abstract

This thesis contributes to a multi-disciplinary research field aiming to build autonomous virtual actors. Emotions are a very active topic of research in many domains, such as biology, sociology, and neurology, due to their powerful influences on human behaviours. Computer science also takes an interest in emotion to create believable autonomous virtual characters. Much research focuses on mechanisms assessing the environment of the virtual actor and generating emotional evaluations to influence the agent's behaviours. In contrast, this thesis looks at the representation of the durations of emotions, the interactions between emotions, and the influences on the visual speech of virtual actors.

Following the review of studies of human facial expression, emotion theories, facial animation systems, and computational models of emotion, a Dynamic Emotion Representation (DER) model and an Emotionally Expressive Facial Animation System (EE-FAS) have been designed. The DER model enables programmers to represent a network of persisting states, where each state can influence the responses of others. An instance of this model is implemented in this thesis and represents three types of state, changing on different timescales. The EE-FAS takes as input a text with tags representing communicative acts and produces visual speech. The presence of the DER in the EE-FAS architecture provides coherence in the production of emotional facial expressions, diversity in the visual representations of communicative acts and variability in the animations produced from a tagged text. The EE-FAS distinguishes and produces emotional expressions due to emotional events and communicative acts synchronised with the speech. To extend the facial vocabulary of virtual actors, the EE-FAS displays different facial expressions corresponding to a single communicative act, according to the DER state. It also distinguishes contexts for displaying genuine and fake facial expressions. An experiment was designed to measure the meanings communicated by facial component actions. This experiment, based on videos produced by the EE-FAS, shows that emotional meanings are perceived from facial component actions, and that the meanings of facial component action combinations are combinations of the meanings communicated by individual components.

Acknowledgements

I am very grateful to Philip Willis for providing me with the opportunity to do a PhD funded by a studentship from the Department of Computer Science of the University of Bath. Many thanks to Phil for his supervision and his support over the duration of my PhD. Many thanks also to Joanna J. Bryson for her enthusiastic support through the ups and downs of a PhD, her social event organisation skills, including the BAI seminars and her financial support during the last year of my PhD through a EPSRC grant (GR/S/79299/01). Thank you to David Duke for his feeback over the first year. Thank you to Leon Watts for his advises on experiment designs.

Thank you to all the PhD students and research officers of the lab for their help, coffee breaks, curry evenings (Bengal Brasserie), BBQs (University residences), pub discussions (The Old Green Tree) and Resurrection nights (Moles). Special thanks to: Sirapat for sharing the PhD experience together from beginning to end; Adam D. who knows how to make a PhD lab a lot more lively; and Marc A. for the long chats over strong cups of coffee. Thank you to all the people from the office for being always friendly and very helpful. Thanks also to Mark and Jim for their computing wisdom.

Thanks to Abi for her supporting love of the grumpy old me. Thank you to my mother, my father and André for their financial and moral support.

Contents

Chapter 1

Introduction

1.1 Emotions: the Art of Communication

Virtual embodied agents or characters become more and more common in the multimedia world of computer games, e-business on the Internet, educative software and films. Many techniques, with different levels of human intervention, exist to enable these characters to move in a human-like fashion. For instance, key frame interpolation methods provide the animators with a precise level of control over the movements of the character, where autonomous embodied agents living in virtual worlds move and act with little human control. The aim of these characters is either to entertain or to interact with people, assuming that at least people understand them. To make this communication more effective, embodied agents simulate the same communicative channels as humans, such as voice, hearing, body movements and facial expressions, even if the agent is a puppy dog or a paperclip.

One particularity of humans is to have emotions; emotions are driving forces motivating our actions. They are so powerful that they push people either to achieve deeds of valour or to behave in the most cruel fashion. Outside of these extremes, emotions incline people's actions in one way or in another. During communication, people assess each other's emotional states, probably because these states are a good indication of what the person could do next, how the person is about to act. Emotions do not show which particular action the person is about to take, but what type of action the person could take. They represent an internal context of the person, a context in which the person's thoughts are processed. In terms of evolution, it might not make sense to reveal emotions from an individual's point of view; but as a social group, humans might benefit from the display of emotions to build and keep relationships and to reject profiteers from social circles.

This thesis is concerned with the communication of virtual embodied agents, in particular with the communication through visual speech, that is facial expressions. Human faces are the most communicative part of the body and display a lot of information in support of dialogue. It is recognised that a link exists between emotions and facial expressions but the intricacy of the relationship between emotions and facial expressions is not well known.

This thesis argues that virtual agents need an emotional state to improve their communication with humans. The selection of facial expressions according to their emotional state would produce a more powerful and a more reliable communicative channel. The work presented in this thesis uses synthetic emotions to select facial signals corresponding to communicative functions, in contrast with other work which use emotions to only produce emotional expressions. In this work emotions are used as a dynamic context providing a relative consistency over time for the selection of facial expressions and potentially other actions. The developed facial animation system also compares the internal emotional state of the virtual actor with the desired emotional context that it intends to communicate, to distinguish contexts for "genuine" and "fake" facial expressions.

Emotions are the heart of human behaviours and play an important role in human communications: virtual agents need emotions to master the art of communication.

1.2 Definition of the Research Area

The focus of this research is the facial animation of virtual actors, because the face is the most communicative part of the human body. The aim of this research is to build a facial animation system with a level of control adapted for either non professional animators or Artifical Intelligence programs, such as autonomous virtual agents. The term "high level of control" does not stand for precise control. On the contrary, high indicates the user has less precision in the creation of animations. The higher the level of control, the less specialist knowledge the users need, or at least the knowledge needed by the users should be more intuitive.

From the review of previous work on facial animation, this thesis defines four increasing levels of control:

1. *Static Graphical Representation*, such as 2D, 3D mesh or volumetric representations. To create an animation at this Level, key postures need to be defined by modifying the graphical representation, such as moving the vertices of a 3D mesh. Then an interpolation mechanism is used to produce smooth transitions between these key postures.

2. *Graphical Deformation Commands or Animation Parameters*, such as Bézier patches, Abstract Muscle models, Facial Animation Markup Languages and the Facial Action Parameters defined by MPEG-4 (Parke and Waters, 1996; Hartung et al., 1998; Pirker and Krenn, 2002; Gustavsson et al., 2001). This Level enables animators to create animations in a more intuitive fashion by stepping away from the graphical representation and using commands related to physical features. For instance, Abstract Muscles enable animators to move several vertices by modifying the value of one Abstract Muscle contraction.

3. *Facial Movements or Facial Signals* are commands at Level of Control 3. They are based on the same descriptions as the previous LoCs but, in addition, functions of time are used to describe the changes of parameter intensity over time. Facial signals describe movements of facial parts and no interpolation mechanism is needed to create animations. Generally these movements are defined over three periods of time: attack, sustain and decay duration (DeCarlos et al., 2002; Kalra et al., 1991; The Duy Bui, 2004; Pelachaud and Bilvi, 2003).

4. *Facial Meanings* are commands at Level of Control 4. These are not physical descriptions of the face but instead they describe meanings expressed

through the face. Examples at this LoC are the Affective Presentation Markup language (AMPL) (Pelachaud and Bilvi, 2003), the high level of the Facial Animation Parameters (FAPs) defined by the MPEG-4 standard (Hartung et al., 1998; Pirker and Krenn, 2002), the Gesture Markup Language (GML) and the Emotion Markup Language (EML) (Gustavsson et al., 2001). Facial meanings can be emotional, such as anger, or communicative acts, such as word emphasising.

Among the four levels of control described previously, this thesis is only concerned with research involving the levels of control 4 and 3. Despite this focus of research, all the levels of control are implemented in the facial animation system designed for this thesis. The user of the facial animation system controls the creation of facial animations using the meanings of facial expressions.

The animations produced by the system should be *believable*, which means that they should mimic human behaviours even if the face is not *realistic*. The visual representation of a face could look like a real face or could be a cartoon, in both cases, the animation is believable if the facial movements looks natural. The facial movements should be interpretable by humans in a way that they communicate the right information, in this case the right facial meaning.

1.3 Definition of the Research Issues

The first problem met when designing a facial animation system with a high level of user control is the limitation of the number of facial expressions that the system can produce. The facial expressions produced from commands at a high level of control are combinations of predefined facial expressions. This limit reduces the communicative power of the face, "the human face can generated around 50,000 distinct facial expressions, which correspond to about 30 semantic distinctions" (Paradiso, 2002, p. 1). So if the system uses the meanings of facial expression as a user interface how can it be possible to produce 50,000 facial expressions?

To extend the number of facial expressions produced by the system, the number of combinations and the number of predefined facial expressions need to be increased. This does not solve the problem because even if it is possible to produce thousands facial expressions using combinations of predefined facial expressions, *how does the system select one of these facial expressions using 30 semantic commands?*

This thesis argues that facial expressions are produced by two different processes: reactions to emotional events and generation of communicative acts. However, it also suggests that the visual representations of communicative acts, called signals, are modified by the current emotional state of the person. The technical solution proposed in this thesis is to use a Dynamic Emotion Representation (DER) to differentiate contexts in which facial meanings take place. The Dynamic Emotion Representation is used to select facial signals corresponding to communicative acts and also to produce emotional expressions, such as Ekman's universally recognised expressions. It represents the emotional state of the virtual actor.and is used to

A number of computational emotion models exist and the question is *what type of emotion representation would be needed for a facial animation system?* This is the second research issue explored in this thesis.

1.4 Contributions

The first contribution is the design of a Dynamic Emotion Representation model enabling programmers to describe relationships between several variable intensities using dynamic filters. This model can be used to describe networks of interacting variable states. In this thesis an instance of this model has been developed to represent three interacting types of states: behaviour activations, emotions and moods. This instance is based on emotion theories and can be used as a mechanism to bias the cognition, the action section, or the expressions of virtual actors. The bias provided by the three types of states can be classified in degrees of emergency: behaviour activations represent states that need to be dealt with immediately, emotions are the context for the current matters, and moods influence all matters over a longer period but with less intensity. The DER composed of these three types of states is used in the Emotionally Expressive Facial Animation System developed for this research to modify the expressions of virtual actors.

The second contribution is the design of a facial animation system linking emotional state and communicative acts. The Emotionally Expressive Facial Animation System (EE-FAS pronounced e-face) uses the DER to:

- control the display of Ekman's emotional full-face expressions as reactions to emotional events;
- select during the speech different facial signals corresponding to a communicative function;
- to produce facial expressions of blend emotions and facial expressions masking emotions.

Each type of state in the DER has a particular function in the EE-FAS: behaviour activations trigger the display of emotional expressions, emotions are used as contexts to select facial signals, and moods influence the general response of the DER. The benefits of a such system are the production of:

- animations dependent on the emotion history of the virtual actor.
- a wide range of animations from a script;
- scripted and unscripted animations;
- emotional expressions coherent over time;
- communicative signals consistent over an emotional episode.

EE-FAS is based on a modular architecture implemented with a message passing mechanism for the communication between modules. Different modules take care of the transformations of commands from one level of control to another. The relation between the emotions, facial meanings and facial expressions are customisable using XML dictionaries, making the EE-FAS a tool for experiment.

1.5 Structure of the Thesis

This thesis is composed of three parts.

The first part reviews the literature in two domains: one about studies of human facial expressions and emotion theories, and another one describing work on facial animation and computational models of emotion. The first literature review is used as a foundation to understand previous work on facial animation and emotion models; and also to develop the Dynamic Emotion Representation model and the Emotionally Expressive Facial Animation System. The facial animation systems and emotion models reviewed represent the state of this art at the moment of the writing of this thesis.

The second part describes the design and implementation of the DER, of the Emotionally Expressive Facial Animation System, and the integration of a Dynamic Emotion Representation within the EE-FAS.

The last part presents and analyses an experiment about the influences of facial component actions on people's perception of virtual actors. The results of this experiment are used to validate the user controls of the EE-FAS. Due to the integration of the Dynamic Emotion Representation in the EE-FAS, the animations produced by the EE-FAS should be judged as having the same emotional characteristics as those used to produce them.

Finally the last chapter presents the conclusions of this research and possible extentions of this work.

Part I

Facial Expressions and Emotions Background

Chapter 2

Background on Facial Expression and Emotion Studies

2.1 Introduction

The aim of this research is to build a facial animation system producing believable animations, using high level commands. High level commands should enable non-specialist animators or computer programs to create facial animations. One of the suggested methods is to provide the users with commands describing the meanings that the face should communicate (Pelachaud and Bilvi, 2003), instead of physical descriptions of facial postures. To explore this route, this chapter presents an overview of work carried out on human facial expressions. The meanings of facial expressions can be explained either by being reflections of internal states, or communicative acts supporting dialogues. As a result of this review, it becomes clear from both positions that a link exists between facial expressions and emotions. This link is not well defined by the research community, so to try to understand the relation between facial expressions and emotions the second part of this chapter introduces the main emotion theories.

By covering these two background domains; studies of facial expressions and emotion theories, this Chapter hopes to provide the necessary information to understand previous work on facial animation and emotions synthesis, reviewed in

Chapter 3. The work presented in the following sections is also used as foundations for the research described in this thesis.

2.2 Studies of Human Facial Expressions

People interpret facial expressions to extract information about the speaker and about the state of the dialogue, but the actual information contained in facial expressions is the subject of a continuing debate.

Based on biological and behavioural needs, Fox et al. (2000) suggest that the mechanism of interpreting facial displays is "hard wired" in the human brain. In the same vein, Fridlund supports the idea that facial expression reading has been selected by evolution; interpretation and display of facial expressions have to "co-evolve" (Fridlund, 1997). In this view, facial expressions are used as a communicative link between people, an idea supported by many researchers (Russel and Fernández-Dols, 1997; Frijda, 1986; Ekman, 1999b; Chovil, 1992; Bavelas and Chovil, 2000)

2.2.1 Facial Action Coding System

One of the major contributions to the analysis of facial expressions is the Facial Action Coding System (FACS) developed by Ekman and Friesen (1978). This system is based on Action Units describing visible movements on the face generated by the contraction of a muscle or a group of muscles. It has been developed to be a standard method to code facial movements from images or videos but now it is widely used in computer animation to describe facial postures. The FACS is very useful for describing facial appearances but it does not provide any information about the meanings communicated through facial expressions.

2.2.2 Facial Vocabulary

To understand what facial expressions communicate, researchers try to break down facial movements into a "minimum vocabulary", a set of basic facial movements carrying meanings. Smith and Scott (1997) summarise three approaches describing how facial expressions communicate meanings: the *purely categorical model*, the *componential model* and the *purely componential model*.

The *purely categorical model* describes a limited set of full-face patterns. In this model, meanings can only be expressed through a full-face configuration, and even if the facial patterns could be described by its components using FACS, each component does not have any meaning by itself. These full-face configurations are often associated with the universally recognised emotional facial expressions. Darwin (1979), Tomkins (1980), Ekman (1992), and Izard (1971) are the main promoters of the existence of facial expressions reflecting particular emotions. Ekman (1992), and Izard (1971) conducted studies across different cultures. They showed pictures of full-face configurations, supposedly typical of particular emotions, and asked the subjects to associate each picture with an emotion word. From these studies, a limited set of emotional facial expressions is qualified as universally recognised facial expressions. Izard (1997) notes that it is not clear what is universally recognised about the emotional facial expressions: it could be positions of individual facial parts or full-face configurations. Two main problems exist with the purely categorical approach: the first one is the limited set of possible meanings produced by the face, which does not match the recognised communicative power of human faces (Smith and Scott, 1997). The second problem is the affirmation made by Russell that full-face configurations do not occur often in normal circumstances (Russell, 1997).

The second model is the *componential model* suggested by Smith and Scott (1997). In this approach, meaningful units are facial component actions, which are related to AUs described by FACS. For instance, an eyebrow frown can communicate anger, frustrastion or an order. The *componential model* suggests that the meaning of the whole might be different than the meaning of the sum of its parts (Smith and Scott, 1997).

In the third approach, the *purely componential model*, meaningful units are also facial component actions but in contrast with the *componential model*, it suggests that the meaning of the full facial pattern is equal to the sum of the meanings of its components.

2.2.3 Facial Expressions as Reflections of Mental States

If different sets of basic facial movements carrying meanings exist, it could be due to the fact that different types of facial expressions exist. Looking at the literature, two main theories explain the production of facial expressions. The first one sees them as a reflection of mental states, which is presented in this sub-section, and the second one, discussed in sub-section 2.2.4, considers facial patterns as *Communicative Acts.*

For a very long time facial expressions have been seen as a reflection of mental states and the most common association is between facial patterns and emotions. Collier (1885), Ekman and Friesen (1969) and Smith and Scott (1997) argue that people look at the face to find signals of emotional states.

Russell (1997) suggests that people extract information from the face of other people through "quasi-physical features", such as gaze direction, how much the eyes are open, and through two dimensions: pleasure and arousal. He declares that, using these two dimensions, people interpret the emotional state of the displayer, but he also emphasises that the emotional state is just an interpretation of the observer and not a reflection of an existing internal state.

Darwin (1979), Tomkins (1980), Ekman (1992), and Izard (1971) believe that emotions produce typical facial patterns. Ekman (1992) associates six universally-recognised facial expressions with six discreet basic emotions: anger, sadness, fear, surprise, joy and disgust. Izard (1997), Ekman (1999b) and Smith and Scott (1997) define a loose link between facial displays and emotions: certain facial patterns are shown during certain emotional states but the presence of emotions is not necessary for these facial patterns to be shown. Also the presence of emotions is not sufficient to produce the related facial displays. Other researchers defined the same type of loose link but instead of talking about emotions as internal states of the displayer, Frijda and Tcherkassof (1997) talk about action readiness, for which emotions are a subset, and Fridlund (1997) talks about intentions and goals. It is worth noting that in fact facial expressions co-occurring with emotions are not exactly the same as those displayed when the emotion is not present. This is the difference between a fake and genuine smile. In the genuine case the eyes "smile" at the same time as the rest of the face, as defined by Duchenne (Ekman, 1992).

Smith and Scott (1997), taking the componential approach of facial expressions, suggest that movements of facial parts, such as eyebrows, could carry meanings. Regrouping work from several researchers, Smith and Scott (1997) propose the following appraisal dimensions defining internal states related to emotions: "...subjective pleasantness of an emotional state, including related appraisals, such as the perception of goal obstacles or the anticipation of the need to expend efforts ...", "...attentional activity associated with emotional state, including the novelty of the situation and one's degree of certainty about one's circumstances ...", level of activation or arousal, and personal agency or control. Each one of these dimensions is expressed through movements of particular facial parts, thus reflecting internal states. Other researchers use different appraisal dimensions to represent emotional state and link these dimensions directly to emotional facial expressions (Scherer, 1988). For instance, *sudden novelty* is related to the Action Units 1, 2, 5, and 26/27, where *intrinsic pleasantness* related to the taste is link to the Action Units 6, 12, and 25/26 (Wehrle et al., 2000).

2.2.4 Facial Expressions as Communicative Functions

A different approach to explain the types of information communicated by facial expressions is to look at their use during conversations (Ekman, 1979; Chovil, 1992, 1997; Pelachaud and Poggi, 1998; Poggi and Pelachaud, 2000; Bavelas and Chovil, 2000). People's dialogues are supported by facial movements, such as eyebrow raising, adding both new or redundant information (Chovil, 1992, 1997). These facial signals are tightly synchronised with the flow of words and change over time too quickly to represent emotional state variations (Ginsburg, 1997; Bavelas and Chovil, 2000).

Ekman (1979) categorises facial expressions into the following groups:

- Manipulator: facial expressions filling biological needs of the face, such as blinking;
- Regulators: facial expressions used to control the dialogue, such as signals giving speech turns;

- Conversational signals: facial expressions used to emphasise words or sentences, or to punctuate the speech;
- Punctuators: facial expressions displayed during speech pauses;
- Emblems: facial expressions replacing the speech and culturally dependent;
- Emotional emblems: facial expressions communicating non-felt emotions;
- Affective displays: facial expressions communicating felt emotions.

Bavelas and Chovil (2000) suggest the term *Communicative Acts* which are composed of a *communicative function*, syntactic or semantic, and its facial physical/visual representation, called the *signal.* For instance, *Portrayal* is a semantic function displayed when the speaker reproduces a facial expressions such as the facial expression of somebody else. Bavelas and Chovil (1997) define the *emphasis* of a word in a sentence as a syntactic function, but it is classified as a semantic function by other researchers. More examples are shown in Table 2.1. The expressions of emotions are part of a sub-set of a wide range of communicative functions which is classified as *personal reaction.* Bavelas and Chovil (1997) argue that facial expressions corresponding to *personal reaction* are different from the typical emotional full-face displays because they could be partial, symbolic, acted, and fast. It is important to notice that the description of facial expressions through communicative functions is not sufficient to define facial signals. As emphasised by Bavelas and Chovil (2000), several signals could correspond to one communicative function and one signal could be used for several communicative functions. This shows the importance of taking context into consideration when interpreting the meanings of facial expressions, as claimed by Izard (1997), Frijda and Tcherkassof (1997) and Fernández-Dols and Ruiz-Belda (1997). Chovil suggests "three aspects of context — the dialogue that has gone before, the other symbolic actions that co-occur with it, and the larger act of which the display is a component" (Ginsburg, 1997, p. 364).

De Rosis et al. (2003) and Pelachaud and Bilvi (2003) give a different description of *Communicative Acts.* In this description, Communicative Acts are also composed of two parts: a *Communicative Function* and a *Signal.* The description of Communicative Acts in Chovil (1997) is close to the linguistic analysis whereas

Table 2.1: Functions of conversational facial action (Bavelas and Chovil, 1997, p. 342, Table 15.1)

Meaningful Displays by speakers			
	Semantic		
		Redundant	
			Personal reaction
			Portrayal
			Thinking
		Nonredundant	
			Personal reaction
			Portrayal
	Syntactic		
		Grammatical markers	
			Emphasiser
			Question marker

the definitions of communicative functions presented by De Rosis et al. (2003) are related to: *belief*, *intention*, *affective state*, and *meta-cognitive*. Each of these four communicative functions are subdivided into classes. Examples are shown in Table 2.2. The *performative class* is interesting because performatives are subgoals of the three main goals, differentiated using social and mental contexts. The main goals are: *request* another agent to do something, *inform* another about a belief, and *ask* another agent to provide its belief about something (Pelachaud and Poggi, 1998). For instance, a *performative* of the type *request* could be an *order*, an *advice* or an *imploration*. The distinction between different performatives is made using the following five characteristics representing the context in which the conversation is situated (Pelachaud and Poggi, 1998):

- *In whose interest* is the answer to a question such as "whose goal does the requested action serve?",
- *Degree of certainty* is used to differentiate between a claim and a suggestion,
- *Power relationship* between the sender and the addressee could be used to make the difference between commands, advises and implorations,
- *Type of social encounter* is used to differentiate between formal and informal talks , and

Table 2.2: Some examples of expressions and their representation with gaze (De Rosis et al., 2003, p. 7, Table 1)

Communicative functions	**Classes**	**Example Meanings**	**Signals**
Belief	deictic adjectival certainty	this, that, here small, subtle big, long, large certain uncertain	Gaze direction Small eye aperture Large eye aperture Small frown Raise eyebrow
Intention	performative topic/comment rhetorical relation turn allocation	order suggest comment contrast Give turn	Frown, chin up, head straight Raise eyebrow, head aside Raise eyebrow Raise eyebrow Gaze direction
Affective state	emotion	joy sadness	Smile, raise cheek Inner raise eyebrow, look down, corner of lip down
Metacognitive	thinking activity	I am thinking	Gaze direction

- *Affective state.*

2.2.5 Recapitulation

Facial expressions can be described at different levels, at a physical level using the Facial Action Coding System, for instance, or at a meaning level.

Two main views exist regarding how meanings are communicated through facial expressions: The *purely categorical* approach suggesting that the meanings of facial expressions are represented by the full-face configurations, and that facial part movements do not communicate any meaning by themselves. The *componential approach* promotes the view that facial part movements carry meanings and that the meaning of a facial expression is a combination of the meanings of its parts.

The meanings of facial expressions can be explained either by the view that facial expressions are reflections of internal states, mainly emotional states, or by considering facial displays as communicative acts supporting dialogues. The point of view taken in this thesis is that both views are valid and that facial expressions are produced by two different process: emotional and communicative processes. However, emotional states are also visible during communicative acts and this thesis argues that they even drive their visual representations.

Universally recognised facial expressions are reflections of discreet emotional states and based on the method used to study them, most people regard them as purely categorical. In these studies, the pictures presented to the subjects were full-face configurations so the immediate consequences were to argue that it is the full-face configurations that have been recognised as products of emotional states (Ekman, 1992; Izard, 1971). From Smith and Scott's componential approach of facial expressions (Smith and Scott, 1997), Izard (1997) suggests that it could be components of facial expressions that are recognised as reflections of emotional states, instead of full-face configurations. Smith and Scott (1997) also believe that facial expressions are produced by emotional state but their description of emotional states is different from the discreet theory of emotions. They use appraisal dimensions, described in section 2.3.3, to described emotional states, and they associate each dimension with movements of particular facial parts.

When facial expressions are seen as communicative acts, they are categorised into syntactic or semantic classes (Chovil, 1997), or as communicating meanings about belief, intentions, affective state and meta-cognitive state (De Rosis et al., 2003). Within this view, a componential approach is adopted.

Sometimes emotions are described as discreet categories with their associated full-face expressions (Ekman, 1992), sometimes as a list of dimensions expressed through movements of facial parts (Smith and Scott, 1997), or sometimes as a subclass of communicative functions (Chovil, 1997; De Rosis et al., 2003). Whether facial expressions are seen as reflections of internal states or as communicative acts, emotional states are definitely expressed through them. The link between emotions and facial expressions is not clear, and as it is discussed in the following section, this is due to the fact that the term *emotions* is not well defined. To throw some light on this discussion, the next section describes the main concepts about emotions.

2.3 Concepts of Emotions

This section reviews the main theories of emotion to try to understand which phenomena are covered by the term *emotion* and what the links between emotions and facial expressions are.

2.3.1 What are Emotions?

Emotion is a common concept. Everybody knows what emotions are, or at least has an idea about what they are. People use vocabulary related to emotions or emotional states every day and generally their interlocutors understand these terms. So why does the research community still not agree on one definition?

> Everyone knows what an emotion is; until asked to give a definition.
> (Fehr and Russell, 1984, p.464)

Frijda (2000, p. 60) lists the following terms which appear in the different definitions of emotions:

> "feelings, shifts in the control of behaviour and thought, involuntary and impulsive behaviours (including "expressive" behaviours), the emergence or tenacity of beliefs changes in the relationship with the environment, and physiological changes not caused by physical conditions."

To explain what emotions are, Descartes (1649) uses an analogy to body pains. Pains are signals alerting that something important is happening to your body; in the same way emotions are signals alerting that something needs attention in our "soul", in our thought (Oatley and Jenkins, 1996). Descartes also emphasises that emotions are phenomena that are not under the control of our mind but it is possible to "recollect" our thought to not let the "soul" be driven completely by passions (Descartes, 1649).

The definition of emotion changes with the point of view taken by the researcher. The points of view of emotions changes across disciplines; emotions are the topic of research in various domains such as philosophy, psychology, neurology, sociology, biology and so on. It is not clear what exactly the phenomena that should be called emotions are; neither is it clear at which level the phenomena should be described (Frijda, 2000).

The number of observation methods used to study emotions are as diverse as the number of fields studying the phenomena. Emotions can be studied through physiological changes, such as heart beat or skin conductivity, through neural activity, or through changes in behaviour (Frijda, 2000).

The observations of physiological changes are the foundations of a theory known as the *James-Lange* theory. James and Lange both independently developed this theory around 1880, defining emotional experiences as "awareness of body responses" to an event, and emotions are produced by body changes such as facial expressions (Oatley and Jenkins, 1996; Frijda, 2000; Edwards, 1999). The James-Lange theory brought forward the importance of the body in the phenomena called emotions. Supporting this theory, Ekman (1992), Tomkins (1980), and Izard (1993) suggest that facial expressions generate, amplify, and/or regulate emotional experiences.

Where the James-Lange theory focuses on the relation between body changes and emotions, researchers such as Darwin (1979), Plutchik (1980), Izard (1977), Tomkins (1980), and Ekman (1999a) study behavioural changes. They defined a set of emotions having a special status called *Basic*, or *Fundamental Emotions*, based on behavioural observations. This type of emotion is generally assumed to be innate and to have particular *survival* functions. The term *Basic Emotions* and the different meanings of the word *basic* will be described in more detail in section 2.3.2.

Emotions are also described as states resulting from, or processes involving cognitive or non-cognitive evaluation of internal and external events in relation to one's well being. This evaluation is known as *appraisal* (Lazarus, 1991; Plutchik, 1980;

Izard, 1993; Frijda, 1986; Arnold, 1960; Ortony et al., 1988; Roseman et al., 1996; Smith and Scott, 1997). Within this theory, many points of view exist, mainly depending on the definition of cognition adopted, but also on the dimensions used to evaluate events. The appraisal theory of emotions and its different definitions are presented in section 2.3.3.

Some researchers, such as Sloman and Izard, suggest emotion models unifying the several definitions of emotion. These models include several mechanisms from which emotions emerge, and define the interactions between these mechanisms. Sloman proposes an architecture-based concept, called CogAff, to study mental processes. CogAff is a multi-level processing architecture schema, and H-CogAff, an instance of this schema composed of three layers, represents human mind architecture (Sloman, 2001b). Within H-CogAff, Sloman describes emotions in terms of side effects of interactions between mental processes within a layer, and also as results of conflicts between processes of different layers trying to get control of the same resources (Sloman, 2000, 2001b,a). The three layers of H-CogAff are called reactive, deliberative and meta-management layers, and each one of them is a source of a type of emotion called primary, secondary and tertiary emotions, respectively. Izard (1993) describes a system integrating four sub-systems, namely neural, sensorimotor, affective and cognitive, each one eliciting emotions. Izard's work is based on a definition of cognition specifying the limits between several types of information processing. Both of these works are presented in more detail in section 2.3.4.

2.3.2 What are Basic Emotions?

The notion of Basic Emotions is probably most known due to its association with universally recognised facial expressions, but this is only one side of the story. The debate within the framework of basic emotions is concerned with questions such as: How many emotions are basic? Which emotions are basic? No general concensus has been reached to answer these questions. The number of basic emotions suggested in the literature varies from 2 to 11 and the overlaps between the different sets of basic emotions are small (Ortony and Turner, 1990; Edwards, 1999).

Ortony and Turner (1990), in "What's Basic About Basic Emotions?", give a wide overview of the theory of basic emotions and discuss the meanings of the word "basic". Their article does not support the existence of a set of basic emotions but clarifies the different opinions and describes the two main definitions of basic emotions. The first definition views the basic emotions as "biologically primitives" and the second definition views them as "psychologically primitives". These two views can be compatible (Ortony and Turner, 1990).

Biologically Primitive

Biologically, basic emotions are innate and serve survival functions developed through evolution. One consequence of this argument is that basic emotions should be found across human cultures and across species of higher animals. The most common method to study basic emotions is the observation of facial expressions (Darwin, 1979; Plutchik, 1980; Izard, 1977; Tomkins, 1980; Ekman, 1999a). The main arguments in favour of the existence of basic emotions are studies carried on by Ekman (1992) and Izard (1971), showing that facial expressions of basic emotions are universally recognised. Ekman (1999a) insists that the only existing emotions are the basic ones. A second consequence to view basic emotions as biologically primitives is that "one would except to find neurophysiological or anatomical evidence of them in all (normal) members of the species" but at the moment this evidence has not yet been found (Ortony and Turner, 1990, page 320).

Psychologically Primitives

The psychological point of view of basic emotions aims to build a set of "psychologically irreducible" emotions, which means that these emotions would be building blocks for other emotions, and that they could not be composed of other emotions (Ortony and Turner, 1990). Within this view, basic emotions can be differentiated by appraisal conditions (Arnold, 1960) or by action readiness (Frijda, 1986).

2.3.3 Emotions in Appraisal Theory

A Simple Definition of Appraisal

The appraisal theory of emotion first started with Arnold (1960) who suggested that the difference between an emotional and non-emotional event is due to the process of appraisal (Parkinson, 1995). This process evaluates an event as beneficial or harmful, as good or bad, or as promoting or obstructing, with regard to one's concerns, motives or goals (Lazarus, 1991; Ortony et al., 1988; Plutchik, 1980; Scherer, 1988). In this literature, one could notice that appraisal and cognition are closely related, the main difference is that cognition is a general information processing mechanism, where appraisal focuses on the evaluation of events. Izard notes that

> "there are many different types of cognition: automatic as contrasted with deliberate . . . ; conscious as contrasted with unconscious, preconscious, and subconscious . . . ; cognition at different levels of awareness . . . ; and cognition in repression and dissociation . . . " (Izard, 1993, p. 70).

Appraisal suffers from the same problem as cognition regarding its definition.

Relation Between Appraisal and Emotions

What is the relation between appraisal and emotions? The answer to this question depends in the first place on the details of the definition of appraisal. Appraisal is an evaluation of an event but it depends on the point of view of the researchers: appraisal can be "basic sensory information processing" (Ekman and Davidson, 1994), can involve fast automatic cognitive processes, or relatively time consuming cognitive processes.

From a biology point of view, LeDoux and Panksepp define cognition as taking place in the neo-cortex and hippocampus. From their studies of patients having important lesions in these brain regions and showing emotional reactions, they conclude that emotions are generated independently from cognition (Ekman and

Davidson, 1994). In this frame of mind the appraisal is defined as "basic sensory information processing" and emotions emerge from this process.

Zajonc defines automatic pre-cognitive processing that could generate emotions on the basis that emotion can be generated faster than in the time needed for cognitive processes (Oatley and Jenkins, 1996). In this case emotions could emerge without cognitive processes. This view, also shared by Izard (Ekman and Davidson, 1994), takes the same definition of appraisal as the biologists.

Most of the people supporting the appraisal theory, such as G. Clore, P. Ellsworth, N. Frijda, C. Izard, R. Lazarus, J. Ledoux, J. Pankseep and K. Scherer, agree that the appraisal generating emotions can be carried out by fast automatic unconscious mechanisms (Ekman and Davidson, 1994; Frijda, 2000). The debate is about including these mechanisms in the cognitive mechanisms or not. As well as these automatic processes, a number of people, such as Plutchik (1980), Lazarus (1991), Frijda (1986), and Izard (1993), support the existence of more complex mechanisms eliciting emotions and involving an agent's memories, belief and goals.

Appraisal Dimensions for Emotion Categorisation

Within the appraisal theory, people use appraisal dimensions to categorise emotions. To find the relation between appraisal dimensions and emotions, subjects are asked to recall an emotional event and to rate the appraisal conditions corresponding to the felt emotion. The appraisal dimensions used in this type of exercise are variable in number and quality. Ellsworth and Smith suggest the nine following dimensions: pleasantness, anticipated effort, attentional activity, certainty, human agency, situation control, perceived obstacle, importance, predictability (Oatley and Jenkins, 1996); Scherer (1988) defines novelty, intrinsic pleasantness, goal/need conduciveness, coping potential, norm-compatibility and self-compatibility; while Roseman et al. (1996) suggest only six: unexpectedness, motivational state and situational state, probability, control potential, problem type and agency; and Smith and Scott (1997) propose seven: pleasantness, goal obstacle or goal discrepancy, anticipated effort, attentional activity, certainty, novelty and personal agency or control. This method of describing emotions is also called the componential approach to emotions (Frijda, 2000). This approach is related to the componential approach to facial expressions through the links between appraisal dimensions and movements of certain facial parts (Smith and Scott, 1997).

Emotions as Appraisal Processes

Plutchik (1980), Lazarus (1991) and Frijda (1986) see emotions as processes in which appraisal is a part. They suggest different steps within this process.

Plutchik (1980) suggests a "chain of emotional reactions":

$$\text{Stimulus event} \Rightarrow \text{Cognition} \Rightarrow \text{Feeling} \Rightarrow \text{Behaviour} \Rightarrow \text{Effect.}$$

For Plutchik, the cognitive process interprets a stimulus event to generate a feeling, e.g. an emotion. This process could be automatic or not and "unconscious" or not.

Lazarus (1991) proposes a process involving two types of appraisal and one of re-appraisal:

- *Primary appraisal* is concerned with the evaluation of an event regarding one's well being. This appraisal includes three components: goal relevance, e.g *Is the event relevant to the current goal?*, goal congruence or incongruence, e.g. *Is the event promoting the goal?* and type of ego-involvement, e.g. *Is the event relevant to self-esteem and social-esteem, moral values, ego-ideals, ... ?*.
- *Secondary Appraisal* is concerned with a coping option: "What, if anything, can I do in this encounter, and how will what I do and what is going to happen affect my well-being?" (p. 134). This appraisal includes blame or credit, coping potential, and future expectations.
- *Re-appraisal* is an appraisal that evaluates the same event but later on, to take into consideration the changes in primary and secondary appraisal,

changes in the environment and one's own actions and reactions.

Frijda (1986) describes the following emotional processes, which take a "stimulus event" as an input and categorises emotions in terms of "action readiness".

$$\begin{array}{l} \textit{Stimulus event} \Rightarrow \textit{Analyser} \Rightarrow \textit{Comparator} \Rightarrow \textit{Diagnoser} \Rightarrow \textit{Evaluator} \Rightarrow \\ \textit{Action proposer} \Rightarrow \left\{ \begin{array}{l} \textit{Physiological change generators} \\ \textit{Actor} \end{array} \right. \end{array}$$

- *Analyser* tries to categorise the input event to a known event type, which has causes and effects associated with it.
- *Comparator* determines the relevance of an event in regard to one's concerns. Frijda calls this relevance evaluation or primary appraisal.
- *Diagnoser* analyses what the possible actions are to cope with the situation. This process is called "context evaluation" or secondary appraisal.
- *Evaluator* uses the information provided by the comparator and the diagnoser to evaluate the "urgency, difficulty and seriousness" of the situation.
- *Action proposer* changes the action readiness. Action readiness determines action tendency, e.g. determine action priorities, reinforces on going action or imposes new actions, e.g. "control precedence".
- *Physiological change generators* reflects physiologically the action readiness state.
- *Actor* selects an action in regard with the action readiness.

Each one of these steps is influenced by and influences "regulator processes". The regulation is concerned with movement control, inhibition, and voluntary self-control. On certain occasions, when the system needs a quick reaction to an event and does not have the time to analyse precisely the situation, some of these steps, such as the Diagnoser, are by-passed. Frijda suggests that emotions can be differentiated using their particular action readiness pattern.

2.3.4 Models with Multiple Emotion Elicitors

From the previous sub-sections, it is clear that the processes producing emotions vary widely depending on the theory adopted. For instance, James-Lange theory argues that emotions result from physiological changes, where people supporting appraisal theory mainly see emotions as a product of cognitive processes. Even within the latter theory the multiple definitions of cognition have an important impact on the decision as to what is producing emotions.

This sub-section presents two theories trying to unify the different views of emotions by defining several interacting mechanisms eliciting emotions.

First, the "Four Systems for Emotion Activation" developed by Izard (1993) is described and then the architecture-based concept of emotions called "CogAff" and designed by Sloman (2003) is presented.

Four Systems for Emotion Activation

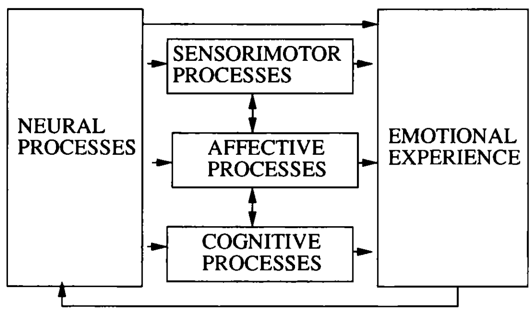

Figure 2-1: A multisystem model of emotion activation (Reproduction of Figure 2 from Izard (1993, p. 74).

Izard (1993), with his paper called "Four Systems for Emotion Activation: Cognitive and Noncognitive Processes", defined a hierarchical organisation of *neural*, *sensorimotor*, *motivational* and *cognitive* systems as sources of emotion elicitation. The bases of this hierarchy are the *neural systems*, which are able to produce emotions by themselves. In these systems, the variation of hormones and neurotransmitters produce emotions. The second set of systems in this hierarchy are the *sensorimotor systems*, body movements, such as facial expressions, regulate, amplify or even produce emotions. Third in Izard's organisation are the *motivational systems*, including drives and emotions. Drives, such as hunger and tiredness can produce emotions and also emotions themselves can elicit other emotions. Finally, the highest in this hierarchy are the *cognitive systems* producing emotions in relation to beliefs and goals. As introduced in subsection 2.3.3, cognition has many definitions; Izard gives the following one:

> "Cognition is about knowledge — learning, memory, symbol manipulation, thinking, and language" (Izard, 1993, p. 73).

The interaction between the systems is represented Figure 2-1. All sources of emotions are based on the *neural processes.*

CogAff Architecture

The difficulty of finding one definition of emotions might come from the possibility that different types of emotions exist. For instance, LeDoux (1999) distinguishes two emotional brain circuits, one being the thalamus-amygdala circuit, which processes rapidly an event stimulus, and the other one being the circuit thalamus-to-cortex-to-amygdala, processing more detailed information (Eysenck and Keane, 2002). In the same sense, Damasio (1994) and Sloman (2001b) define *primary*, and *secondary* emotions and Sloman also describes *tertiary* emotions. Sloman bases his description of emotion types on a theory dividing the brain structure into three layers. Each one of these layers hosts a type of mechanism, namely *reactive*, *deliberative* and *meta-management* mechanisms. Within this view, different types of emotions are generated by each one of these mechanisms and by their interactions. Sloman also suggests that this brain structure has been developed through evolution. In these terms, the reactive mechanisms are the oldest and the meta-management mechanisms are the newest. This view is important because the presence of the oldest parts of the brain structure, such as the reactive mechanisms, in both humans and animals, could explain why they also share the same type of emotions. Sloman also suggests that emotions such as shame or guilt are typically human and are produced by the third layer where the meta-management mechanisms take place (Sloman, 2001b, 2003). Below is a description of the different mechanisms and the types of emotions that they produce.

- *Primary emotions* are produced by reactive mechanisms mapping external stimulus patterns to behaviours. These mechanisms are qualified as "fast and dirty" because they enable fast reactions to certain recognised stimulus patterns but these behaviours could be wrongly activated. In fact, reactive mechanisms involve little if any cognitive processes, so little context is taken into consideration. This sometimes results in the recognition of stimulus patterns and the triggering of behaviours in inappropriate situations. Primary emotions are innate rules mapping stimulus patterns to pre-organised behaviours (Damasio, 1994). The definition of primary emotions is very similar to the definition of basic emotions given by Darwin (1979), Plutchik (1980), Izard (1977), Tomkins (1980) and Ekman (1999a). In evolutionary terms, the cost of this type of mechanism is the amount of space needed to memorise the rules, pushing for the development of more general purpose mechanisms, e.g. deliberative mechanisms, and giving birth to secondary emotions (Sloman, 2001b, 2003).

- *Secondary emotions* are learned associations between recognised stimulus patterns generating primary emotions and analysed situations where these patterns occurred (Damasio, 1994). The processes involved in learning and analysis of situations is called a deliberative mechanism by Sloman. Deliberative mechanisms are cognitive processes taking into consideration goals, beliefs, standards and expectations, enabling reasoning about situations, plan making, and understanding of action consequences. They are also responsible for creating hypothesis and abstract concepts (Sloman, 2001b). An example of a secondary emotion is hope, which is not due to the recognition of external stimulus patterns but due to a mental picture of possible future events. If deliberative mechanisms are more adaptive than reactive mechanisms to new situations due to the analysis of situations and plan making, they are also slower due the involvement of memories and cognitive processes, so a trade-off needs to be found between reactive and deliberative control of the behaviour.

- *Tertiary emotions* emerge from conflicts over limited resources between deliberative and meta-management mechanisms. Meta-management mechanisms enable the cognitive awareness of internal processes or states and provide the possibility to reason about these internal states and processes. Sloman argues that the interaction and competition for resource control between declarative and meta-management mechanisms could result in complex internal states typical of human emotions. Meta-management mechanisms are also called self-monitoring mechanisms. Sloman suggests that the redirection of attention or the loss of control of the meta-management mechanisms are typically tertiary emotions, but he also admits not having a clear idea of what mechanisms produce this (Sloman, 2003, p. 106). One could imagine that the meta-management mechanisms never have control of the thought processes but instead these mechanisms, running in parallel, overview the general state of the natural system and redirect the attention to issues that need to be dealt with.

2.3.5 Emotions vs Mood

When researchers talk about affect, they talk mainly about emotions but other terms come into discussion, such as temperament, feeling, and personality. The second most discussed affective term after emotions is mood. Researchers suggest two main differences between emotions and mood, which are their object directedness, or the lack of it, and their durations.

Frijda (1986), Lazarus (1991), Oatley and Jenkins (1996), Sloman (2003), and others (Ekman and Davidson, 1994) suggest that emotions, as described in previous sections, are object oriented. For instance, when somebody is angry she/he is angry about something, about somebody. On the other hand, mood is objectless, or mood takes anything and everything as an object (Sizer, 2000). When somebody is irritable, she/he could be irritated by anything and everything. Sizer (2000) argues that mood should be seen as a "cognitive functional architecture", which is a list of rules regulating how the information is learned and recalled from memory, how the information is categorised and how other resources are used. This argument is supported by a number of experiments, such those carried out by Schare and Bower (Oatley and Jenkins, 1996). Sizer (2000) argues that mood is "cognitively impenetrable", meaning that mood is not influenced directly by cognitive processes; instead, it represents an overview of the person's internal state. This also means that mood is not directly influenced by particular beliefs or situations but by the changes in the overall state of a person.

The second difference between emotions and mood is their durations. Emotions can last from under a minute to a few minutes whereas mood can last for hours, days or even weeks (Oatley and Jenkins, 1996; Sloman, 2003; Ekman and Davidson, 1994; Thayer, 1996). Ekman (1999a) argues that one characteristic distinguishing emotions from other affective states is that emotions have particular recognisable facial expressions associated with them. Ekman and Davidson (1994) suggests that mood might be reflected through non-visible facial muscle tonus. The tonus of a muscle is a state of partial contraction.

Thayer (1996) describes mood using two "arousal continuums": energy to tiredness and tense to calm. Tense arousal appears in stressful or demanding situations; the person is on the alert, nervous or anxious, ready for action, if she/he can. Energy arousal is a subjective energy that includes mental energy and energy depending on physical resources. Thayer (1996) bases his description of energy on the argument that the body and mind cannot be separated. In Thayer's description, mood can be seen as a self-monitoring mechanism in the way that mood is a representation of a person's state, physical and mental, and it influences cognition in relation to this mood state.

2.3.6 Recapitulation

Research on affect phenomena has a long history and is still very active. The term *emotion* has many definitions, mainly due to the fact that this term includes several types of phenomena depending on the point of view of the researchers who study it. Emotions are not palpable, they can only be observed through physiological, behavioural, and neurological changes. The definitions of emotions also depend on the definitions of other terms such as *cognition*.

Two models are based on multiple mechanisms eliciting emotions, those described by Sloman (2003) and Izard (1993). Where Sloman's model is based on a mind architecture schema, Izard takes a more biological approach, taking into consideration hormones and neurotransmitters.

This section also introduced the term mood, emphasising the differences between emotions and mood. It is worth noticing the resemblance between the definition of "meta-management" mechanisms given by Sloman (2003) and the definition of mood given by Thayer (1996) and Sizer (2000). All these definitions include a description of mechanisms "over looking" and controlling other mechanisms. These resemblances are discussed in Chapter 4.

This Chapter shows that most of the research regarding emotions focuses on the mechanisms eliciting emotions. The differentiation between emotion types is mainly based on the mechanisms producing them. The principal information about the duration of an emotion is the length of time for which its corresponding facial expressions last. Even the difference between primary and secondary emotions is not about their durations but about the time taken by the processes eliciting them. One interesting distinction between mood and other emotions is its longer duration.

2.4 Conclusion

This Chapter presented two different, but related background literatures: studies of facial expressions and of emotion theories. The human face, with its expressions, is often seen as a communicative channel but the content of this communication is the subject of a continuing debate.

The content of the messages communicated by the face can be studied on the one hand by the point of view that facial expressions are reflections of internal states, such as emotions; and on the other hand facial expressions can be seen as communicative acts supporting face-to-face dialogues. One overlap between these two views is the communication of emotions through facial expressions.

The second part of this chapter presented the main emotion theories. It is necessary to review emotion theories to try to understand the relation between facial expressions and emotions. Many emotion theories exist, changing according to the research area of the author, such as psychology, neurology, philosophy, sociology or biology. The most unifying emotion theories are those suggesting the existence of multiple systems eliciting emotions; several interacting mechanisms producing emotions. As far as the relation between facial expressions and emotions is concerned, a little information is known. Universally recognised facial expressions are linked to basic emotions and they might be displayed at the activations of basic emotions Izard (1997). Another view is to see universally recognised facial expressions as accumulative movements of facial part from a sequential appraisal of an event Wehrle et al. (2000). Facial expressions related to communicative acts can communicate emotional messages but their visual representations might also be influenced by the emotional state of the person.

The background presented in this chapter is also necessary to understand the computational work carried out on facial animation and emotion synthesis, presented in Chapter 3.

Chapter 3

Background on Facial Animation and Emotion Models

3.1 Introduction

The previous Chapter presented several theories about facial expression categorisations and gave an overview of emotion theories. This Chapter describes how these categorisations and theories are applied into computational systems. It reviews existing systems using a high level of control to create facial animations and emotion models used in autonomous virtual actors. This review compares the characteristics of these systems and emphasises the place of the work presented in this thesis.

The first part of this Chapter, Section 3.2 describes the different techniques used by different systems to create believable facial animations. Most of these systems produce visual speech using emotional expressions or communicative acts. The second part, Section 3.2.6, explains why virtual actors need emotion models and afterwards it reviews computational emotion models classified by the types of emotion elicited. At the end of this Section, emotion models are compared to each other and to the Dynamic Emotion Representation developed for this thesis.

3.2 Facial Animation System: Talking Heads

3.2.1 Modification of Facial Signals

EMOTE and FacEMOTE presented by Badler et al. (2002) are interesting solutions for changing the expressiveness of a character. These solutions are based on Laban Movement Analysis (LMA) which is a method "of movement study for observing, describing, notating and interpreting human movement" (Badler et al., 2002, p. 136). Using the components Effort and Shape of the LMA, EMOTE modifies neutral movements to express the character's mental state, such as its emotional state. As emphasised in the paper, EMOTE is not an action selection mechanism, it is used later on in the process to influence selected movements or facial signals.

3.2.2 Facial Animation from Text Analysis

BEAT, described by Cassell et al. (2001), takes pure text as input to generate "embodied expressive behaviours". The source text is tagged by the system with linguistic and contextual information, such as theme and rheme, action and new object. This information is used to suggest nonverbal behaviours, such as gesture beat, gesture iconic and eyebrows. Over a complete utterance, a list of signals is suggested for each piece of added information. Signals with low priority and creating conflicts with other signals are filtered. Conflicts are described over a list of degrees of freedom.

Albrecht (2005) describes a facial animation system taking a text as an input. By analysing the text, the system produces emotional synthetic speech and synchronised facial movements. Each facial movement could be from one of the five following types: emblems, illustrators, affect displays, regulators and adaptors. These facial signals selected according to the information provided by the "Mary Text-to-Speech". For instance, head and eyebrow movements are randomly displayed at local pitch maximum, and eye-blinks appear at speech pauses. This system uses an emotion model, based on the "emotional wheel" described by Plutchik (1980). In this model, the emotional space is represented by a disk defined by two dimensions: activation and evaluation. Six basic emotions, namely joy, anger, fear, hate, sorry-for, and surprise, are mapped onto the emotional wheel and their typical facial expressions are used to produce facial expressions. When an emotional word is found in the text, it is mapped onto the emotional wheel specifying the emotional expression that should be produced. If the emotional word corresponds to a basic emotion, its related emotional facial expression is displayed, otherwise a combination of the two closest basic facial expressions is shown. Similarity between two emotions is proportional to the angle that separates their positions on the wheel. Albrecht (2005) describes an algorithm to combine two facial expressions based on a bounded weighted sum of the muscle intensities of the two combined facial expressions. Albrecht (2005) describes a solution to resolve physical conflict when facial expressions are combined. Physical conflicts are due to the combination of facial movements that are not physically possible. To detect physical conflicts a set of antagonistic muscles is created for each muscle. The contraction intensities of the muscle involved in a conflict are inversely proportional to the importance of the conflict. The importance of the conflict is computed by the difference between the contraction intensities of the muscles involved in the conflict, before they were influenced by the conflict. Although physical conflicts have been taken into consideration by this algorithm, conflicts at a semantic level have not been dealt with.

3.2.3 Goal Orientated Communicative Functions

Pelachaud and Bilvi (2003) present a facial animation system taking a text tagged with Affective Presentation Markup Language (APML) as input. APML describes *communicative functions* that need to be expressed in synchronism with words present in a text (Poggi and Pelachaud, 2000; Carolis et al., 2002). The notable advantage of augmenting a text with communicative functions is that it adds information about what should be expressed without specifying how it should be expressed (Pelachaud and Bilvi, 2003). This means that the signal, e.g. the physical realisation of a communicative function, could be different in relation to many characteristics of the virtual character, such as its body, its emotional state or its personality (Pelachaud and Bilvi, 2003). De Rosis et al. (2003) describes four communicative functions: *Belief*, *Intention*, *Affective state* and

Metacognitive. Each communicative function has a list of classes, for instance, *Belief* has *Deictic*, *Adjectival*, and *Certainty* classes or *Affective state* which has only one class called *Emotion*. Several communicative function tags can cover the same part of the text, which means that several facial movements should be displayed at the same time. If these movements involve the same facial part a conflict can occur. The example given by Pelachaud and Bilvi (2003) presents two communicative functions, a *performative order* and a *comment* which are realised as a *frown* and a *raise eyebrow*, respectively. As pointed out in this paper, "adding these two signals would not produce a believable expression" or at least would communicate the wrong meaning (Pelachaud and Bilvi, 2003, p. 18). The solution proposed in this paper is to choose one of the two signals using a belief network. One problem with this solution is the loss of communicative information by the fact that all the communicative signals might not be shown. Another solution could be to have several possible physical realisations for each communicative function, and to use some kind of consistency parameter to select the appropriate, e.g. non-conflicting, signals. For instance, in the previous example the signal corresponding to the communicative function *comment* could be a more intense *eyebrow frown* than the one related to the communicative function *order*. It should be emphasised that in this system, emotions are only used as communicative functions and do not influence the selection of signals corresponding to other communicative functions. The use of emotional states to select signals corresponding to communicative functions is one of the improvement suggested in this thesis.

3.2.4 Multimodal Face-To-Face Architecture

The architecture presented by Thórisson (1999), called Ymir, is used for creating autonomous virtual agents capable of "fluid multimodal dialogue". With respect to the domain of research of this thesis, Ymir implements some interesting features. Gandalf, which is an agent based on the architecture Ymir, communicates using his face, through *emotional emblems* and *communicative signals*, his hands, his eyes, his body, and his speech. The description of movements is carried out at three levels: high level such as a *smile*, medium level such as *Pull-Corner-of-Mouth-Up* and low level such as *Move-Motor-x-to-Position-y*. The higher the description level the more physical realisations exist. For instance, many ways of smiling exist, where fewer ways exist to pull the lip corners up (Thórisson, 1999). Each high level description, i.e. each behaviour, has a list of lower level descriptions enabling virtual agents to express behaviour with different signals. Signals corresponding to a behaviour are scored according to the number of conflicts they produce with the current running behaviour and the signal with the lowest score is selected to be active. This selection mechanism is based on the physical state of the virtual character, where this thesis suggests the use of emotional states to select signals.

3.2.5 Emotion Representations and Emotional Expressions

The emotional wheel model described by Plutchik (1980) is used by several facial animation systems, including Kurlander et al. (1996), Raouzaiou et al. (2003), and Latta et al. (2002). This popular model enables researchers to build mechanisms to map emotional state to facial expressions but it is a static emotion representation. This type of representation does not provide any consistency mechanism for the production of emotional expressions. Emotions change slowly and a minimum of time should separate the expressions of two opposite emotions such as anger and happiness. With a static emotion representation any expression can be displayed independently from the previous facial expression. The following systems integrate Dynamic Emotion Representations, which is the type of emotion representation that is used in the facial animation system developed for this thesis. Dynamic Emotion Representations keep track of the changes in emotion intensities, representing an emotional momentum, which provides a consistency mechanism for producing emotional expressions.

The work described by Kshirsagar and Magnenat-Thalmann (2002) uses an emotional model to select facial expressions. A Bayesian Belief Network represents personality and mood state and it is used to select the next emotional facial expression that the character will display. As input, the system takes a list of possible responses associated with an emotional state probability provided by a chat-robot system called ALICE. For each emotional state probability, the system computes the probability of each possible mood state in relation to the personality of the character and the previous emotion and mood state. The mood state with the higher probability is selected and used to choose one of the emotional states suggested and its associated text. The character can be in one of 24 emotional states, which are mapped to one of the six universally recognised facial expressions by having the same expression for multiple emotional states. The interesting characteristic of this system is the use of previous states to select the next emotional expression. However, the major drawback is the limited number of facial expressions that can be displayed by the system. Their emotion model will be described in more detail in section 3.3.

In the architecture of the agent called Obie, described by The Duy Bui (2004), six basic emotions, namely fear, surprise, anger, disgust, sadness and happiness, are used to generate emotional facial expressions. Two fuzzy rule-based systems map an emotional state vector to facial expressions: one to map a single emotion to its universally recognised facial expression, and one to map the two emotions with higher intensities to a blend of their universally recognised facial expressions. A module called *expression mode selection* "chooses" if one or two emotions should be displayed. If the difference between the two higher emotion intensities, ranging from 0 to 1, is greater than 0.5 only one emotion is expressed. The blending of the facial expressions, defined by a set of rules, is carried out by expressing the two emotions in different parts of the face. This mechanism solves the issue of conflict between facial meanings because only one meaning is expressed by one part of the face. One interesting point of using fuzzy rule-based systems to map emotions to facial expressions is the possibility to have different muscles involved in the expression of an emotion depending on the intensity of this emotion. For instance the system could produce the differences between a small smile with closed mouth due to a low intensity of happiness and a smile with open mouth due to a high intensity of happiness. As an extention, The Duy Bui (2004) suggests that the *expression mode selection* could be influenced by personality parameters to select which emotions should be displayed and which emotions should not be displayed. The Duy Bui (2004) also suggests that the character could modify the emotion state vector (ESV) by changing the intensity of emotions, making the ESV different from the "felt" emotions, to produce fake emotional facial expressions. One problem with this solution is that the fake facial expressions will be identical to a genuine facial expression, making the character a very good liar. The emotion model of this system is described in section 3.3. The Duy Bui (2004) describes an algorithm to resolve physical conflicts of facial movements. This method is based on looking forward into the animation to stop potential conflict.

3.2.6 Comparison of the Reviewed Work

This thesis argues that emotional expressions due to emotional events and represented by the universally recognised expressions are different than facial expression due to communicative functions, even if they can communicate emotions too. Table 4.1 compares the systems reviewed in this section. The comparison is based on which types of facial expression the systems display, what kind of emotional representation, if any, is used and the criterion used to select facial signals. The first part of Table 4.1 provides the author names and an index that should be used in the second part of the table.

3.3 Computational Emotion Models

This section focuses on existing emotion models and their functions in a virtual agent architecture. Most of the work reviewed is not particularly concerned with facial animations but it is interesting to see in an overview how emotion models are used in a broad agent architecture and what the different emotion models are.

3.3.1 Why Might Virtual Characters Need Emotions?

Why Might Virtual Characters Need Emotions? To answer this question, the functions of emotions in natural systems should be considered.

Table 3.1: Comparison between facial animation systems

	Author
A	Badler et al. (2002)
B	Cassell et al. (2001)
C	Pelachaud and Bilvi (2003)
D	Albrecht et al. (2005)
E	Kshirsagar and Magnenat-Thalmann (2002)
F	The Duy Bui (2004)
G	Thórisson (1999)

	Type of Expressions Displayed	**Emotion Representation**	**Criterion to Select Facial signals**
A	facial signals	none	n.a.
B	comm. functions	none	minimisation of conflicts
C	comm. functions	none	none
D	emotional expressions and comm. function	static	voice pitch
E	emotional expressions	dynamic	n.a.
F	emotional expressions	dynamic	n.a.
G	comm. functions	none	minimisation of conflicts
EE-FAS	emotional expressions and comm. functions	dynamic	emotional state

Functions of Emotions in Natural Systems

Most researchers agree that emotions have functions (Cañamero, 1998), from the definition of Descartes (1649) who defines emotions as an alarm system about something important occurring in the thought process, to the appraisal theory of emotions which describes emotions as evaluations of events from the agent's well-being point of view, through to the definitions of basic emotions which attribute to them survival functions.

> "The biological function of emotions is twofold. The first function is the production of a specific reaction to the inducing situation. ... The second biological function of emotion is the regulation of the internal state of the organism such that it can be prepared for the specific reaction." (Damasio, 1999, p. 53–54)

> "Emotions serve something, and presumably they serve it well. They serve concern satisfaction; they do so by monitoring the relevance of events and by modulating or instigating action accordingly." (Frijda, 1986, p. 475)

Some of the functions of emotions in a natural system are:

- "Recruiting physiological changes" to prepare the agent to act, this includes for instance the increase of the blood flow (Ekman and Davidson, 1994; Damasio, 1999).
- Activation of innate behaviours which have survival function, such as jumping away from danger (Izard, 1977; Darwin, 1979; Plutchik, 1980; Tomkins, 1980; Damasio, 1994; Ekman and Davidson, 1994; Ekman, 1999a).
- Changing the cognitive process by gathering information emotionally relevant to the situation and focusing attention (Eysenck and Keane, 2002; Ekman and Davidson, 1994; Ortony, 2001).
- Communicating intentions and motives to other agents (Frijda, 1986; Ekman and Davidson, 1994; Ortony, 2001).

Why Might Virtual Characters Need Emotions?

As pointed out by Cañamero (1998), there are two reasons to "give" emotions to a virtual character. The first reason is to use emotion-like mechanisms in an agent architecture to reproduce the beneficial functions that emotions have in natural systems. These types of mechanisms simulate what happens in a natural system. The second reason is to use emotion-like mechanisms to exhibit emotional behaviours of natural systems, without simulating the complex processes involved in emotion elicitation and the effects of emotions on cognitive processes. It is the same distinction as that between the simulation-based and communication-driven approaches described by Gratch et al. (2002). Gratch et al. (2002) distinguish communication-driven virtual agents selecting emotional expressions that need to be communicated to observers, and simulation-based virtual agents simulating the processes eliciting emotions and expressing the emotions that are "felt", not selected to be communicated. These two types of agent are not incompatible, they can be seen as two sub-systems of a virtual agent architecture.

Cañamero (2003) is concerned with the influence of emotions on action selection mechanisms in virtual agents. An action selection mechanism is in charge of selecting what to do next, which action should be taken next. Cañamero (2003) describes three functions of emotions regarding action selection: "Body adaptation" enabling a quick reaction to particular situations, "Motivating and guiding action" by appraising positively or negatively events and changing goal priorities, and "Expressive and communicative function" which informs other agents of the agent's action tendency.

Ortony et al. (1988) explain that emotions can be seen as a positive or negative appraisal of consequences of events, actions of agents and aspects of objects with regard to the agent's goals.

Cañamero (2003) suggests that the design of emotional agents should be based on the nature of the environment and the tasks of the agent in this environment. The level of complexity of the emotional architecture depends on"What does this particular (type of) agent need emotions for in the concrete environment it inhabits?" (Cañamero, 2003, p. 12).

Following this explanation of why emotion models are integrated into agent architectures, the next sections present several emotion models, starting by the well-known OCC model (Ortony et al., 1988).

3.3.2 The OCC Model

This section is dedicated to a brief presentation of the Ortony Clore Collins' (OCC) model due to its popularity for synthesising emotions cognitively elicited.

Ortony et al. (1988) describe "The Cognitive Structure of Emotions", which has been developed to enable Artificially Intelligent programs to reason about emotions. This structure, also called the OCC model from the names of its authors, is a model of causation because it enables a system to find what type of emotions can occur from the appraisal of "consequences of events", "actions of agents", or "aspect of objects", with regard to the goals, attitudes and standards of the agent concerned. The intensity of an emotion can be calculated using intensity variables such as the desirability of an event, the praiseworthiness of an action and the appeal of an object. In its original version, the OCC model can be used to elicit 22 emotions, including emotions regarding an other agent's concerns such as "happy-for". The consequences of events could be appraised with the focus on the consequences for other agents or for itself, as the appraisal of praiseworthiness of an action could be focused on the other agent or on itself.

The first version of the OCC model has been judged too complex to be implemented completely in an emotional virtual agent (Bartneck, 2002). Ortony (2001) reduced its emotional structure to 10 emotions by eliminating all the branches for the concerns of other agents.

The OCC model is popular because it enables researchers to think about emotions in a structured way and to implement mechanisms eliciting emotions from events occurring in the environment of the agent, other agents' actions and objects. It is the only model which provides enough detail to be implemented relatively easily.

A few criticisms have been voiced about the OCC model. Bartneck (2002) notes that the computation of the intensity of emotions should use a history function of past events, in such way that the same event occurring consecutively would not produce the same emotional intensity. The problem with this comment is that the desirability of this event should have changed after its first occurrence, therefore the computed intensity should be different. Bartneck (2002) also discusses the number of emotions elicited through the OCC model on the basis that they cannot all be expressed through facial expressions, limited in his system to the six universally recognised facial expressions. This position does not take into consideration that "emotion response-tendencies" affect not only the expressiveness of the character but also its information processing and coping (Ortony, 2001). Bartneck (2002) and Picard (1997) point out that the OCC model does not inform about the dynamics of emotions and the interactions between emotions. This thesis sees the OCC model as an appraisal tool eliciting potential emotions due to cognitive processes with regard to the goals, standards and attitudes of the agent. The dynamics of emotions and the interactions between emotions should be treated by a separate entity because this mechanism is not relevant to the cognitive processes, neither is it relevant to goals, attitudes and standards. Emotions do affect the cognitive processes, even the cognitive processes generating emotions, simulated by the OCC model, but their influences are due to the changes of goal priorities, attitudes and standard values. Examples in favour of this argument are the systems described by The Duy Bui (2004) and Reilly (1996), which incorporate a module to elicit emotions using the OCC model, and a different module to represent the dynamics of emotions. The Emotionally Expressive Facial Animation System (EE-FAS) developed during this research is based on the same dichotomy.

The OCC model is used in numerous systems to elicit emotions in relation to the agent's goals, belief, and standards. Some of the systems based on the OCC are reviewed in the later sub-sections but first the next section presents systems eliciting emotions in different ways.

3.3.3 Model of Primary Emotions

Delgado-Mata and Aylett (2004) describes a "Neuro-fuzzy Model of the Amygdala" which is used in the implementation of virtual animals such as in virtual sheep or animals inhabiting a virtual antique Mayan civilisation city. These animals are autonomous in a virtual environment, flocking together and communicating emotion to each other through the release of pheromones. The communication of emotions through pheromones influences the behaviours of the other animals of the same species. The amygdala is the part of the brain generating primary emotions and this paper presents a model of the amygdala based on fuzzy logic systems which are implemented by neural networks (Delgado-Mata and Aylett, 2004). The emotions of these virtual animals are generated from five inputs: two internal motives which are well-being and hunger; two external emotion signals produced by other animals which are fear and joy, and a signal from the environment called Flight Zone. If a predator agent is situated in the Flight Zone, the animal would run away to keep a safe distance from the predator. The virtual animals communicate their emotions by releasing virtual pheromones which are received by other animals through a virtual nose. The emotional state of an animal is computed using a hierarchical structure of five Fuzzy Inference Systems and it is represented as the 3-element vector: anger, fear and joy. The emotional state is used to influence the action selection mechanism controlling the virtual animals. In Delgado-Mata (2004), an emotion intensity is computed by adding the decay of this emotion and the sum of all "affecting variables A_j to the emotional system", which can be positive or negative (p. 113). By supposing that the "affective variables" are the five inputs of the emotional model previously described, one would believe that emotions do not influence each other, except if the animal can smell their own released pheromones, which is not specified in either paper.

The agents called Abbots, presented in Cañamero (2003), inhabit a 2D grid world, where they are looking for food and water, running away from the predators and playing with the agents of the same species. Their actions are driven by goals which are selected depending on the values of *motivations*, such as hunger, thirst, aggression, and social bonding. Each motivation is associated with a *controlled variable*. If the value of a controlled variable is out of the viability range the associated motivation becomes active. Controlled variables are influenced by hormones released within the agent when an emotion is triggered. In this system emotions are seen as amplifiers of motivations. Each emotion releases one hormone, which influences several controlled variables. The motivation with the highest value tries to control the behaviour of the agent and the motivation with the second highest value is the source of opportunist behaviours, occasionally taking control of the behaviour. This mechanism enables emotions to change the priorities of goals, for instance in urgent situations, by influencing the values of motivations. Emotions also directly affect physiology, attention, perception and behaviour. Based on the environment characteristics and the agent's possible behaviours, six basic emotions have been implemented: anger, boredom, fear, happiness, interest and sadness. Each emotion can be triggered by particular internal and external stimuli, and has particular functions for the survival of the agent.

3.3.4 OCC Based Models

The Oz project, described in Reilly and Bates (1992), aims to develop interactive, "dramatically interesting", emotional agents situated in a virtual physical world. These agents are based on an architecture called *Tok*, working on a sense-think-act cycle. *Hap*, a subsystem of Tok, is an action selection mechanism driving the agent's behaviour in relation to the active goals and their importance, the previous actions, the emotional state and the personality of the agent. The assessment of the success or the failure of goals by Hap is used by another subsystem, called *Em*, to update the emotional state of the character. Em is both an emotional and a relationship model but this section describes mainly the model of emotions. *Emotion generators*, principally based on the OCC model, produce *emotion structures* of different types. Most of these emotion structures are generated in relation to external events and to the agent's goals. The emotion intensity is computed in relation to the importance of goals. For instance, the failure of a goal would produce an emotion of the type *distress*. Emotions like fear and hope are generated in relation to a pre-condition to the success of a goal. For instance, for the goal "get-human-to-open-door", the agent would hope if there is a human around. These kinds of emotion can be classified as *secondary emotions* but the system can also elicit emotion types that can be qualified as *primary emotions*. For example, the emotion type *startle*, also classified as a *reflex emotion*, is due to the detection of a loud noise in the virtual environment. The character's goals do not intervene in the processes producing this emotion type (Reilly, 1996). An emotion structure contains information about the emotion, such as its intensity, its cause, and its direction (such as being angry towards somebody). Different linear decay functions are associated with each emotion type. Emotion structures of the same type are combined using three types of functions: *winner-takes-all*, *additive*, *logarithmic combination* (Reilly, 1996). Emotions do not directly influence the behaviour of the character; instead *behavioural features* are used to direct behaviours. Behavioural features have the same structure as the emotion structures but they can be modified by the system to change the expressivity of the characters, such as "feigning" emotions, or associating different behaviours to particular emotions in relation to the personality of the character (Reilly, 1996). A broader emotion type, mood, is also computed based on the behavioural features. Moods could be of several types, due to their foundations on behavioural features but in the described implementation mood varies on the dimension bad/good in relation to the number of positive versus negative emotion structures. Emotion structures, through behavioural features, influence behaviour by adding new goals, changing the priorities of the goals, making goals easier or harder to reach or to miss (Reilly, 1996). They also influence body state, including facial expressions. The set of facial expressions is composed of frowning, smiling, scowling, bug-eyed, and several colours for the face, which could be the default colour, flushed or pale. The behavioural feature with the highest intensity is selected to influence the character's behaviours but if no behavioural feature has an intensity superior to a certain threshold, bad or good mood is used for this function (Reilly, 1996). A particularity of Em is that it also represents relationships between agents on the dimension like/dislike (Reilly and Bates, 1992). The state of relationships influence the emotions of the character. For instance, if a disliked agent is around the character, it would become angry. Relationships change over time depending on whether agents help or impede the success of the character's goals (Reilly and Bates, 1992).

EMA, developed by Gratch and Stacy (Gratch, 2000; Gratch and Marsella, 2004b), is an emotion model based on Lazarus and Smith's "cognitive-motivational-emotive" theory described in section 2.3.3 (Gratch and Marsella, 2004b) and implemented using the OCC model. This emotion model focuses on the influences of emotions on cognition processes and behaviour responses of virtual characters. Events are appraised using plan descriptions, belief and intentions to represent causal relationship between events and states; this process relates to the primary appraisal described by Lazarus (1991). The result of the appraisal is summarised by a list of seven appraisal values including perspective, desirability, likelihood and so on. Using these appraisal dimensions, an emotional state, e.g. the name of an emotion type and its intensity, is computed through an implementation of the OCC model (Ortony et al., 1988). Using this emotional state, the appraisal dimensions, memories, planning and perception, a coping strategy is selected among ten choices; this process relates to the secondary appraisal described in Lazarus (1991). The coping strategies could be an action, a plan, or a positive reinterpretation type. The result of the coping process could change intentions, belief or plans, so a re-appraisal might take place following the coping strategy. To have consistency between the movement of different parts of the character's body, one of three "Physical Focus Modes" is selected in relation to the emotional state: "body-focus", "transitional" and "communicative". The Physical Focus Mode is used to select gaze movements, and hand gesture types of the agent but it also influences the perception of events. In "body-focus" mode, a character would filter out events with low priorities. As far as the emotion model is concerned, the focus of this work is the development of a mechanism eliciting emotions and the influences of emotion on cognition and body movements. These two papers do not describe the interactions between emotions, nor the representation of emotion intensity changes.

The Duy Bui (2004) describes a system called ParleE which is built of five components: a *planner* producing "optimal plans" for the agent and computing the probability of achieving goals, *models of other agents* to calculate the probability of events, an *emotion appraisal component* based on the OCC model, an *emotion component* representing and updating emotions, and an *emotion decay component* providing the values of emotion decay. The *emotion appraisal component* generates *Emotion Impulse Vector* (EIV), representing an emotion type generated by the appraisal of an event with regard to the agent's goals, standards, and expectations. In addition to the emotion types described by the OCC model, ParleE takes into account the emotion type *surprise* which is associated with unexpectedness of an event, related to the probability of the event happening. For each emotion type two thresholds are defined: α and ω. When the intensity of an emotion type is higher than α, this emotion type becomes a "felt" emotion, otherwise it is considered to be a mood. Moods have less influence on the agent's behaviour than emotions. The threshold ω is the upper limit of an emotion intensity. The intensities of the emotion types are computed for each goal and the intensities of each identical emotion type are summed. The *emotion component* uses the *Emotion Impulse Vector*, resulting from the appraisal process and including all the emotions having an intensity superior to the threshold α, to modify the *Emotional State Vector* (EVS). The EVS, representing the emotional state of the character, is updated by computing the intensity of each emotion. The intensity of each emotion within the EVS is computed by adding its decay function and the new impulse of this particular emotion multiplied by the sum of the influences of the other emotions on this particular emotion. ParleE also integrates four motivations: hunger, thirst, fatigue and pain; generating goals and driving the behaviour of the agent. Motivations with high values influence the "emotional experience" of the character by modifying the thresholds α. Personality also influences how emotions are computed and how they change. Personality parameters modify α and ω thresholds, values of appraisal variables such as expectation, and emotion decay functions.

3.3.5 Models With Multiple Emotion Types

The system presented by Reilly (1996), can be seen as eliciting two types of emotions, primary and secondary emotions. Startle, a primary emotion, is elicited directly by the reading of external sensors and the recognition of the occurrence of a loud noise. Other emotion elicited through a mechanism based on the OCC model can be viewed as secondary emotions.

The Presence project aims to design an affective agent architecture for a virtual guided tour delivering information and answering questions about the German Centre for Artificial Intelligence (DFKI) (André et al., 1999). Some of the requirements for the agents are to be customisable through personality traits, to generate believable behaviour and to be affectively expressive. The architecture is composed of two channels running in parallel. The reactive channel generates primary emotions enabling the character to react quickly to the user's actions. The deliberative channel generates secondary emotions with regard to the agent's goals, belief, and attitudes using the OCC models, and also organises dialogue plans. Upstream within the system, events are classified and appraised in relation to the agent's goal and then action requests are communicated to the reactive module and dialogue goals are communicated to the declarative module. Emotions and personality influence the choice of the phrases, the involvement of the character in the dialogue and the priorities of the higher goals. The output module implements and synchronises actions required by the reactive and deliberative modules, taking into consideration the emotional state of the character, such as anger or sadness, but it also takes a more general emotional description in terms of arousal and valence. No details about how this module works are provided in this paper.

The work described by Paiva et al. (2004) is concerned with the creation of believable autonomous characters which would generate empathy in the person watching them. This work is part of a project called VICTEC trying to find new solutions to fight bullying at school. In this system called "Fear Not!", the characters should be able to model the state of the world, their own emotional state and those of other characters, and they should be able to express their affective state. Emotions are generated through event appraisal, which is carried out using two OCC models: one to appraise events in relation to the goals that the character actively tries to achieve and another one to appraise events related to goals that the character cannot control. These appraisal processes involve the character's standards and beliefs and they are influenced by the personality of the character. An emotion elicited by one of these processes is "felt" by the character only if its intensity is higher that a certain personality threshold, otherwise the emotion is ignored. Emotion intensities decay over time and when they reach zero the emotion is not "felt" anymore. The action selection mechanism is composed of two layers, schematic and coping layers. The schematic layer activates pre-organised behaviour if an emotion intensity is higher than a certain threshold. This layer simulates primary emotions. The coping layer involves the same coping processes as those presented in (Gratch and Marsella, 2004b) and described in a previous paragraph. The choice of the coping strategy is also influenced by the personality of the character, for instance, a bully has the tendency of carrying on his bullying, and the victim might adopt a strategy of situation reinterpretation (Paiva et al., 2004). The characters are represented as 2D cartoon children and their facial expressions are based cartoon versions of the six universally recognised expressions.

3.3.6 Emotion Representation Models

Cathexis, developed by Velásquez (1997), models emotions and mood dynamics, temperaments and their influences on behaviour. The system is based on a number of "proto-specialists" taking internal and external sensors as inputs and running in parallel. Two types of proto-specialists are implemented: emotions and drives. The sensors of an emotion proto-specialist can be of the four following types: Neural, Sensorimotor, Motivational and Cognitive. These sensors are defined by Izard (1993) as four types of emotion elicitors. The system includes six basic emotion types, therefore six emotion proto-specialists: anger, fear, disgust, sadness, happiness, and surprise. Emotion proto-specialists can inhibit or excite other proto-specialists. An emotion intensity is computed using the "previous level of arousal for that emotion (which takes into account the mood), the contribution of each of the emotion elicitors for that particular emotion, and the interaction with other emotions (inhibitory and excitatory inputs)" (Velásquez, 1997, p. 3). Emotions and mood are differentiated by the lower level of arousal and the longer duration of the mood. Two thresholds, α and ω are defined for each emotion type. α is the intensity threshold over which an emotion is considered as an emotion, in contrast to a mood. ω is the maximal intensity for an emotion type. The temperament of the agent is represented by the variation of the emotion thresholds. Five drive proto-specialists are implemented: hunger, thirst, temperatureRegulation, fatigue and interest. Behaviours are triggers from drives, emotion and external sensors. Behaviours are competitive and a behaviour becomes active if it has the highest value, which is computed using the values of the related drives, emotions and external sensors. The active behaviour has an experiential component modifying emotional state and motivation values, e.g. drive values, and an expressive component which acts on the facial expressions, body posture and non-language vocal expressions. The facial expressions are based on the six universally recognised expressions. This system can be seen as eliciting different types of emotions due to the use of several emotion elicitors.

The work described by Egges et al. (2004) focuses on the representation and the interactions of emotions, in contrast with mechanisms to elicit emotions. This kind of work can be placed on the side of an OCC based system to keep traces of the dynamics of emotions and to simulate emotion interactions. Egges et al. (2004), who carried on the work from Kshirsagar and Magnenat-Thalmann (2002), describe a personality-mood-emotion model where at a time t, the personality p, mood m_t, and emotion state e_t are represented as variable size vectors of float values. A "desired change in emotion intensity" a, coming from an OCC based mechanism for instance, is represented as vector of float values of the same size than the vector representing the emotional state e_t. A new emotional state e_{t+1} is computed as a sum of the previous emotional state e_t, the influence of the "desired change in emotion intensity" a with consideration of the personality p and the emotional history ω_t ($\Psi_e(p, \omega_t, a)$), and the update of the previous emotional state (emotion decay) with consideration of the personality p and the emotional history ω_t ($\Omega(p, \omega_t)$):

$$e_{t+1} = e_t + \Psi_e(p, \omega_t, a) + \Omega_e(p, \omega_t).$$

To take the mood into consideration, a two step process is used to get the emotional state at a time $t+1$; first,the algorithm updates the mood state and then it computes the emotional state. The same type of equations as the one previously described are used to calculate the mood state m_{t+1}:

$$m_{t+1} = m_t + \Psi_m(p, \omega_t, \sigma_t, a) + \Omega_m(p, \omega_t, \sigma_t),$$
where σ_t is the history of mood.

The computation at $t + 1$ of the emotional state affected by the mood is now carried out with the following equation:

$$e_{t+1} = e_t + \Psi_e'(p, \omega_t, \sigma_{t+1}, a) + \Omega_e'(p, \omega_t, \sigma_{t+1}).$$

For more details refer to Egges et al. (2003, 2004). As specified in Egges et al. (2004), the current implementation does not take into consideration the emotion history during the computation of emotion intensity, e.g. $\Psi_e'(p, \omega_t, \sigma_{t+1}, a)$ becomes $\Psi_e'(p, \sigma_{t+1}, a)$. One consequence of not using the emotion history in this computation is that emotions cannot influence each other. For instance, if the character is very angry and something nice is happening to it, the character will stay very angry and become happy too.

3.3.7 Comparison of the Reviewed Work

This thesis defines emotion models as being composed of two parts: the mechanisms eliciting emotion from the virtual agents' environment, such as mechanisms based on the OCC model, and mechanisms representing how the emotion intensities change overtime. It is two different tasks to assess emotional events from the environment and to track the natural emotion intensity fluctuations and interactions between emotions. The link between mechanisms eliciting emotions and emotion representation is emotion structures. Certain emotion structures are vectors of emotion intensities that have been computed by the emotion elicitor mechanisms (The Duy Bui, 2004; Egges et al., 2004). A position in these vectors represents the type of emotion. Some emotion structures represent only one elicited emotion but more information is carried out such as the emotion intensity, emotion type, and the cause of this emotion (Reilly, 1996). These emotion structures modify the current state of the emotion representation.

In this thesis, emotions are used to influence the expressions of a virtual actor therefore it is less important to know how and why the emotions have been elicited than what is the current emotional state of the character. Much work focuses on the elicitation of emotions and many of them have a minimal emotion representation, such as decay of an emotion after its elicitation. The role of emotion representation is not only to compute the decay of emotion intensities, but also to represent how emotion structures influence the state of the emotion representation. For instance, the fact that the virtual actor broke a glass is a negative event, the mechanisms eliciting emotion would produce a negative emotion structure, such as anger, but how this emotion structure influences the emotional state of the virtual actor depends on its current emotional state. If the virtual actor is in a positive emotional state, in a good mood, this event might not have much effect on the emotion representation, whereas if the virtual actor is in a negative emotional state, it might become very angry and lose its temper. Someone could argue that the variation of the value representing the desirability of an event, which is used by the mechanisms eliciting emotions, should produce this difference by changing the intensity of the generated emotion structure. In this case, the desirability of all the possible events should be changed. Instead emotional states can be used to bias reactions to events, without changing their desirability. Emotions are more of a general purpose mechanism amplifying or reducing the relevance of events.

Table 3.2: Author names for the comparison between the emotion models in Table 3.3

	Author
A	Delgado-Mata and Aylett (2004)
B	Cañamero (2003)
C	Reilly (1996)
D	Gratch and Marsella (2004b)
E	The Duy Bui (2004)
F	André et al. (1999)
G	Paiva et al. (2004)
H	Velásquez (1997)
I	Egges et al. (2004)

For these reasons, the comparison between the different computational emotion models concentrates on the emotion representations of these systems. The comparison is presented in Table 3.3, with the names of the authors in Table 3.2. The column *Emotion Elicitor* in Table 3.3 shows what kind of mechanism is used to elicit emotions. The column *Link between elicitor mech. & emo. rep.* presents the type of structure used to link emotion elicitor mechanisms and emotion representations. The third column, *Type of emotion*, specifies what types of emotion, generally classified by the type of emotion elicitor that produces them, are in the emotion representation. The column *Number of emotion* shows which emotions are represented. The column *Intensity computation* describes how the intensity of an emotion is computed after its elicitation. Finally, the column *Emotion interaction* shows what interactions exist between emotions, if any. The row name *D.E.R* in Table 3.3 is the characteristics of the Dynamic Emotion Representation developed for this thesis.

3.4 Conclusion

This Chapter reviewed facial animation systems controlling the face with high level commands. Two main types of facial expression have been distinguished,

Table 3.3: Comparison between the emotion models

	Emotion elicitor	**Link between elicitor mech. & emo. rep.**	**Type of emotion**	**Number of emotions**	**Intensity computation**	**Emotion inter-actions**
A	5 external and internal stimuli	?	Primary emotions	Anger, joy fear	Decay	
B	External and internal stimuli	?	Primary emotions	Anger, happiness, fear, boredom, interest, sadness	constant computation	
C	OCC model and external stimuli	Emotional structure	Primary + secondary emotions	Startle + OCC emo.	Decay	
D	OCC Model appraisal + coping strat.	?	Secondary emotions	OCC emo.	?	
E	OCC Model	Emotion Impulse Vector	Secondary emotions + mood	OCC emo. + surprise + mood	Decay + other emo.	Dynamic emotion threshold
F	OCC Model + external stimuli	?	Primary + secondary emotions + mood?	? + OCC emo. + mood?	?	
G	2 OCC Models + external stimuli appraisal + coping strat.	?	Primary + secondary emotions	? + OCC emo.	Decay	
H	Neural + Sensorimotor + Motivational + Cognitive	?	Primary + secondary emotions + mood	Anger, happiness, disgust, sadness, fear, surprise, + mood	Decay + emotion elicitors + mood	Emotions inhibitory and excitatory
I	Example with OCC Model	Emotion intensity vector	emotion + mood	Any number of emo. + any number of moods	Decay + mood influence	
D. E. R.	Any elicitor mech.	Emotion Impulse (intensity, name, valence)	Example with: primary, secondary emo. + mood	Any number of emo. types, any number of emo. within each emo. type. Example with: 6 behaviour activations + 6 secondary emo. + mood	Attack, sustain, decay	Dynamic emotion threshold + other emo. influences

facial expressions due to emotional events and facial expressions due to communicative functions. This thesis argues that the first type is related to Ekman's full-facial expressions and does not occur during the speech; the second type is synchronised with the speech and involves combinations of facial part movement. Most of the existing systems do not make this distinction and use Ekman's facial expressions to express emotional messages during the speech. Facial expressions related to a communicative function can have different visual representations, i.e.

facial signals. Existing animation systems select a facial signal of a communicative function to reduce the number of conflicts between co-occurring movements or according to the current physical state of the virtual actor. The Emotionally Expressive Facial Animation System developed for this thesis uses the emotional state of the virtual actor represented by a Dynamic Emotion Representation (DER) to select facial signals and to control the display of emotional expressions.

This chapter emphasises that emotional models are composed of two parts: a mechanism eliciting emotions and an emotion representation. It also shows that most work focuses on emotion elicitation. In comparison to other emotion representations, the DER can represent any number of emotion types and it also implements a network of influences between emotions. Due to the definition of emotional impulses as an interface between mechanisms eliciting emotions and the Dynamic Emotion Representation, the latter can be used with different mechanisms. The integration of the DER in the EE-FAS provides a consistency mechanism to select visual representations of communicative functions and to produce emotional expressions due to the relatively slow changes in intensities of emotion over time.

This Chapter and the previous one presented the background of this work and they will be used to understand the research issues and as a guide lines to develop the EE-FAS and the Dynamic Emotion Representation (DER).

Part II

Emotionally Expressive Facial Animation System

Chapter 4

A Dynamic Emotion Representation Model

4.1 Introduction

Due to the relationship between facial expressions and emotions, the Emotionally Expressive Facial Animation System (EE-FAS) integrates a Dynamic Emotion Representation (DER) within its architecture. This Dynamic Emotion Representation keeps track of emotional episodes producing emotional expressions and it is also used as an emotional context, in particular to select visual representations of the communicative acts.

This thesis suggests that the term of *emotion model* should be defined as containing two separate parts:

- a mechanism or mechanisms eliciting emotions from external and internal stimuli, including potentially the agent's own goals, beliefs and standards; and
- the emotion representations keeping track of the emotional states and their changes over time.

In the design of emotion models the distinction between mechanisms eliciting emotions and emotion representations is useful; the assessment of an emotional event can be the same but its impact on the emotional state of the virtual actor can vary according to the current emotional state. For instance the event of knocking over a cup of tea might make somebody already angry lose his temper, whereas if this person was happy in the first place this negative event might just reduce his level of happiness. Emotion representations enable programmers to reduce the complexity of mechanisms eliciting emotions by assessing an identical event in the same way. However, emotion representations can be seen as general mechanisms to bias any assessment of events, what ever the events are.

As has been emphasised in Chapter 2, the duration of emotions and their interactions are not the focus of much research. The same imbalance is found in computational models of emotions, the focus is on the mechanisms eliciting emotions. To be able to work on emotion representations, the assumption is that mechanisms eliciting emotions produce emotional impulses that modify the state of the emotion representations. Emotional impulses can be defined by the name of an emotion and an intensity value.

The Dynamic Emotion Representation model described in this Chapter is novel because:

- it can represent any number of persisting states, such as moods, emotions, drives;
- any emotion impulse can affect any state positively or negatively;
- state responses to emotional impulses are influenced by the other states represented in the DER;
- it is totally customisable through an XML file.

The instance of the DER model integrated in the EE-FAS is also presented in this Chapter. This DER is composed of three types of states changing on different timescales: behaviour activations, emotions and moods. The representation of these states has been guided by emotion theories and the needs of the EE-FAS. It is shown in this Chapter that this DER keeps persisting states and that according to the original state of the DER, emotional impulses modify its state in different ways. Even if the DER has been implemented for the needs of the

EE-FAS, it could also be used within a larger architecture to influence cognitive processes, action selection mechanisms and expressions of virtual actors with different degrees of emergency.

Section 4.2 presents the concepts of Dynamic Emotion Representations and the Dynamic Emotion Representation model. Section 4.3 describes the DER implemented for the EE-FAS and shows some examples how its states change over time.

4.2 A Dynamic Emotion Representation Model

This section presents the concept behind Dynamic Emotion Representations and relates the DER model to existing emotional models and emotion representations. Afterwards, the techniques used to implement the DER model are described.

4.2.1 Concepts of Dynamic Emotion Representations

The concept of Dynamic Emotion Representation comes from the division of emotion models into *mechanisms eliciting emotions* and *emotion representations.* Mechanisms eliciting emotions assess events from the internal and external environments of the virtual actor to give them an emotional value. These mechanisms are based on different appraisal theories (Lazarus, 1991; Ortony et al., 1988; Plutchik, 1980; Frijda, 1986; Izard, 1993) and produce one or several emotional values for each situation. These values are generally used to influence the decision processes and the behaviours of virtual actors. Certain systems contain an emotion representation to maintain states changing over time which is influenced by the emotional values produced by the mechanisms eliciting emotions.

Some work on emotion models focuses on the mechanisms eliciting emotions (Cañamero, 2003; Gratch and Marsella, 2004a; André et al., 1999; Delgado-Mata and Aylett, 2004). Paiva et al. (2004) present an emotion model which assigns a decay function to each emotion elicited with a value higher than a personality threshold. This work does not implement any interaction between emotions in its emotion representation. Egges et al. (2004) describe a generic emotion and personality representation composed of two types of affective states, moods and emotions. Any number of moods and emotions can be represented. Their model proposes that the computation of the next emotion intensity should use: (i) the previous intensities of all emotions and moods, (ii) the intensity of this emotion at the current time step and (iii) the decay in intensity of this emotion. All of these are to be influenced by the current mood state and personality parameters. However, their implementation does not include (i) for the computation of emotion intensities, therefore the only interaction between affective states is the influence of moods on emotions. Egges et al. (2004) computes the intensity of affective states by linear functions through the use of matrix operations. The Duy Bui (2004) also uses a decay function to represent the durations of emotions, but the effects of new emotional impulses on an emotion are influenced by the intensity of the other emotions. Their decay functions are also influenced by personality parameters. In Velásquez's representation, the computation of the intensity changes of an emotion takes into consideration the intensity of other emotions, the decay in intensity and the previous intensity of the emotion itself (Velásquez, 1997). The influences of emotions on others are of the types inhibitory or excitatory.

The work described in this Chapter focuses on Dynamic Emotion Representations. The role of a Dynamic Emotion Representation is to maintain states, in particular emotional states. The term *dynamic* is used to point out that emotions are ongoing, persisting, constantly changing states. This term also emphasises that the influences of the mechanisms eliciting emotions on the emotion representation depends on the original state of the emotion representation. In contrast to existing systems, the DER model designed for this thesis can represent any number of types of states, such as moods, emotions, and drives, which can be composed of any number of components, such as happiness and sadness.

The Dynamic Emotion Representation model assumes that mechanisms eliciting emotions produce *emotional impulses*, which are defined by the name of an emotion, an intensity value and a valence, which can be 1, 0 or -1. The valence specifies whether an emotional impulse is positive, neutral or negative. Emotional impulses can been seen as a simplified version of the *emotional structures* described by Reilly, which also contain the cause of the emotions (Reilly, 1996). In other emotion models, mechanisms eliciting emotions produce vectors of floats representing emotional impulse intensities and the indexes of the vectors are the emotion types (The Duy Bui, 2004; Egges et al., 2004). Emotional impulses are defined as interfaces between mechanisms eliciting emotions and emotion representations. They are produced by mechanisms eliciting emotions and modify the state of dynamic emotion representations.

In contrast with other emotion representation, the DER model enables programmers to customise the influence of emotional impulses. Any emotion impulse can affect positively or negatively any states represented in the DER. For instance, a happiness impulse can increase the level of happiness and the positive value of mood but it can also reduce the level of anger.

The DER model is based on Picard's description of emotion intensity and emotion filters (Picard, 1997). The next sections explain how the DER model is implemented with regard to Picard's descriptions.

4.2.2 Computation of Emotion Intensities

Emotions appear from event evaluations, and their intensity decay slowly after reaching a peak. Picard (1997) uses a parallel between changes in emotion intensity and the sound of a struck bell. The sound of a struck bell can be modelled using the equations: 4.1, 4.2 and 4.3 producing the graphs in Figure 4-2. These equations can be used to represent the fast increase and slow de of emotion intensity after the appraisal of emotional events. The parallel between emotion intensity and the sound of a struck bell is also due to the suggestion that a series of emotional events are additive and as the sound of a bell gets louder after a succession of strokes, the emotion intensity gets higher. Figure 4-2 shows the emotion intensity response due to a series of emotional events, represented by a series of *emotion stimuli*. This type of response is relatively close to the galvanic skin response of a person hearing startle tones (Picard, 1997). Galvanic skin response is one of the measurements used to get some insights on emotional characteristics because it varies over emotional episodes.

$$\begin{aligned}&\text{Attack: } Intensity = a\ln(t+1)\\&\text{For } 0 \leq t \leq D_{attack}\end{aligned} \tag{4.1}$$

$$\begin{aligned}&\text{Sustain: } Intensity = I_{peak}\\&\text{For } D_{attack} \leq t \leq D_{attack} + D_{sustain}\end{aligned} \tag{4.2}$$

$$\begin{aligned}&\text{Decay: } Intensity = b\ e^{-(c*t)}\\&\text{For } D_{attack} + D_{sustain} \leq t \leq D_{attack} + D_{sustain} + D_{decay}\end{aligned} \tag{4.3}$$

where D_{attack} is the attack duration, $D_{sustain}$ is the sustain duration, D_{decay} the decay duration, and I_{peak} is the peak intensity of the emotional stimulus. a, b and c are coefficients related to the duration of the emotional stimulus.

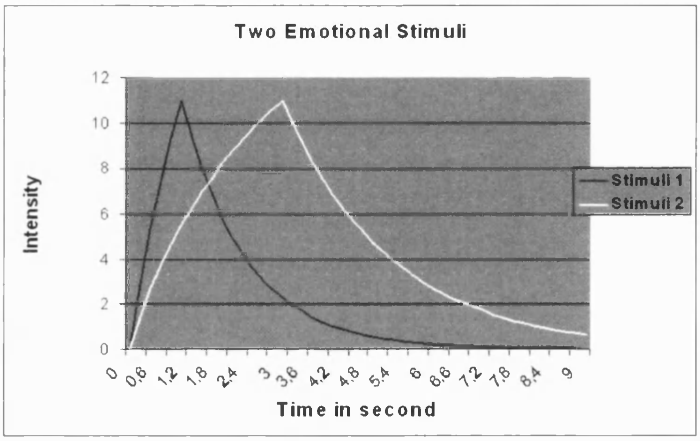

Figure 4-1: Example of two emotional stimuli with different attack and decay duration.

Picard's method to compute emotion intensity is closer to behaviours found in natural systems than linear methods. The characteristics of an emotional stimulus, e.g. its attack, sustain and decay duration, are typical to the type of emotion it represents. For instance, the emotion surprise lasts for a relative short time in comparison to the emotion sadness. This can be represented by fixing different durations for each type of emotional stimulus. A emotional stimulus is defined by its decay, attack and duration, and from these values the coefficient a, b and c need to be computed. From the durations of a particular emotional stimulus,

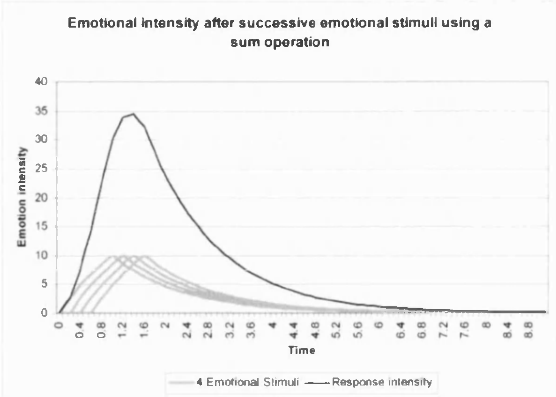

Figure 4-2: Operation *sum* used to compute the intensity of emotions, as suggested by Picard (1997).

a, b and c, in the equations 4.1 and 4.3, are computed with the equations 4.4, 4.5 and 4.6, respectively. These equations come from the equations 4.1 and 4.3.

$$a = \frac{I_{peak}}{\ln(D_{attack} + 1)} \tag{4.4}$$

$$b = I_{peak} \tag{4.5}$$

$$c = \frac{\ln(\frac{Precision}{I_{peak}})}{-D_{decay}} \tag{4.6}$$

where I_{peak} is the peak intensity of the stimulus, D_{attack} is the attack duration, D_{decay} is the decay duration and $Precision$ is the value for which the intensity of the stimulus will be considered as equal to 0.

Emotional impulses defined by an emotion name, an intensity value and a valence, are transformed by the Dynamic Emotion Representation model into emotional stimuli. The name of the emotion is used to select the corresponding attack, sustain and decay durations and the peak intensity is equal to the emotional impulse intensity.

Table 4.1: Parameter values of the two stimuli shown in Figure 4-1

	Attack Dur. in sec.	Sustain Dur. in sec.	Decay Dur. in sec.	Peak Int.	Precis.	a	b	c
Stimulus1	1	0	9	11	0.01	15,87	11	0,78
Stimulus2	3	0	15	11	0.01	7,93	11	0,47

4.2.3 Dynamic Filters

According to her current emotional state, a person will emotionally react differently to certain emotional events. To reproduce this kind of behaviour, the influences of emotional impulses on the DER state should depend on its current state. To implement this characteristic, sigmoid functions are used as *dynamic filters*. This technique is inspired by the description of the influence of mood on sigmoid functions given by Picard (1997). According to the state of a Dynamic Emotion Representation, sigmoid function parameters are modified, resulting, for example, in a shifted sigmoid curve as show in Figure 4-3. A sigmoid function is defined by the equation 4.7.

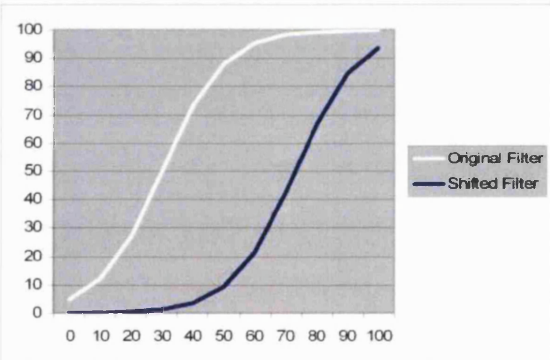

Figure 4-3: Sigmoid functions are used as dynamic filters by changing their parameters.

$$y = \frac{g}{1 + e^{-(x-x_0)/s}} + y_0 \tag{4.7}$$

where s is the steepness of the curve, x_0 shifts the curve left or right, g is the gain and y_0 shifts the curve up or down.

Emotional stimuli, as those shown in Figure 4-1, are modified by these dynamic filters. In most cases, peak intensities of the emotional stimuli are modified. In Figure 4-3, the horizontal axis represents the input value, such as the peak intensity of an emotional stimulus, and the vertical axis represents the output value, such as a new peak intensity value. This mechanism modifies the effects of emotional impulses according to the state of the Dynamic Emotion Representation. Dynamic filters are not only used to modify the peak intensity of emotional stimuli but also other characteristics of emotional stimuli, such as attack, sustain and decay duration. Figure 4-3 shows how the parameter x_0 of a sigmoid curve can be modified to change the intensity of emotion stimuli. The dynamic filter shown in Figure 4-3 can be used to modify the intensity of negative emotional stimuli. The original filter is for neutral mood and it is shifted to the right when the mood is positive. This mechanism reduces the effects of negative emotional events in case of positive mood by reducing their peak intensity. In the same vein, thresholds of emotion activation can be modified simulating different emotional states or personalities.

4.2.4 Description of Emotional Spaces

In the Dynamic Emotion Representation model, emotional spaces, such as moods or emotions, are represented as modules. Each module is composed of a list of dimensions. A dimension represents a variable intensity equal to the sum of emotional stimuli, as described earlier. Each dimension is associated with a group of dynamic filters controlling the effects of emotional stimuli on this dimension.

For example, a module can represent the emotional space described by the *emotional wheel*, which is composed of two dimensions: activation and evaluation (Plutchik, 1980). Another module can represent an emotional space containing all the six basic emotions, one dimension per emotion.

The next section describes an instance of the Dynamic Emotion Representation model composed of three modules.

4.3 A DER Composed of Three Modules

This section presents the DER designed for the EE-FAS. It explains what the three modules of this DER represent and why they have been implemented. Examples on how a series of emotional impulses influences the state of the DER according to its original state are also explained. Finally, an overview of the XML configuration of this DER is described.

4.3.1 Description of the Three Modules

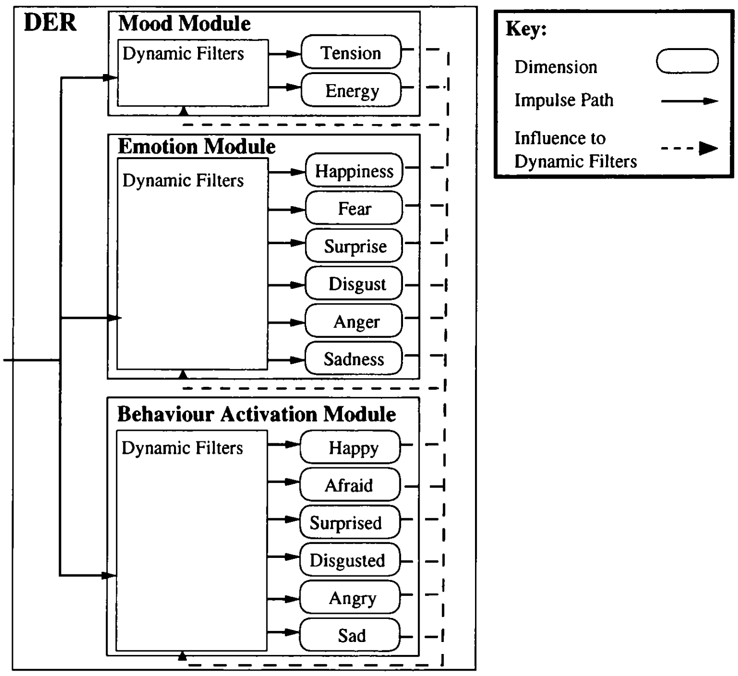

Figure 4-4: The three modules of the dynamic emotion representation.

From the Dynamic Emotion Representation model a DER is created. This DER, integrated into the EE-FAS, is composed of three modules: *behaviour activations*, *emotions* and *moods*. Each represents persisting states changing on different timescales. Now it will be explained why these states have been chosen and how they relate to the production of facial expressions in the EE-FAS. Figure 4-4 shows a graphic representation of the implemented DER.

Behaviour activations come from the definition of *Primary Emotions.* Primary emotions enable natural systems to react quickly to changes in their environment by triggering pre-organised innate behaviours at the detection of particular patterns (Damasio, 1994; Sloman, 2003). A dimension in the first module stands for a behaviour activation, triggered by particular types of emotional impulses. Due to the representation of behaviour activations in the DER, it is possible to control the thresholds and intensities of the activations of the behaviours. According to the state of the DER, thresholds and intensities are modified, making the behaviours activations dependent of the emotional state and the emotional history. For instance, a person who is scared might have a different reaction to a close loud noise than an angry person. Primary emotions can be implemented in mechanisms eliciting emotions to enable virtual actors to react quickly to particular conditions in the environment.

There is no explicit link between pre-organised innate behaviours triggered by primary emotions and facial expressions. However a link exists between Basic Emotions and facial expressions. Basic emotions are defined as being the psychological and biological basis of emotions found across humans (Ortony and Turner, 1990). Basic emotions are defined as being innate and their existence is supported by the recognition of six emotional facial expressions across several cultures, one expression for each basic emotion (Ekman, 1992; Izard, 1971). Izard proposes that these facial expressions might be displayed at the *emotion activation* (Russel and Fernández-Dols, 1997). Universally recognised expressions are considered as full-face configurations due to the method used to study them. This suggests that universally recognised facial expressions can be considered as pre-organised innate behaviours activated when an basic emotion is elicited. Inspired by this suggestion, the first module of this DER is composed of six behaviour activations, namely Happy, Afraid, Surprised, Disgusted, Angry and Sad. For each behaviour activation, the EE-FAS displays one of the six universally recognised facial expressions. Each behaviour activation can be triggered by an emotional impulse of the right type and lasts for under a second. This mechanism enables the EE-FAS to display clear emotional messages due changes in the environment through the effects of emotional impulses on the DER.

Here, the term *Emotions* indicates emotional states lasting for under a minute to few minutes. Emotions and moods are represented in this DER.

If emotional impulses can produce behaviour activations, they can also produce more lasting emotional states, which are represented in the second module of the EE-FAS DER. It is not clear which emotions exist in natural systems. As basic emotions are the most common categorisation, we included the six basic emotions as lasting states in the second module. In the EE-FAS, emotions are used as contexts in which communicative functions take place. According to the state represented by the second module, a different facial signal corresponding to a communicative function is selected. This mechanism enables the EE-FAS to display different facial expressions according to the DER state, as further explained the next chapter.

Looking at theories of emotions, a more lasting emotional context is mood; it is generally differentiated from other emotions due to this characteristic. As suggested by Picard (1997), mood could influence the effects of emotional stimuli on other emotions. For instance, "A bad mood can make it easier for negative-valenced emotion to be activated, while a good mood makes this more difficult" (Picard, 1997, p. 155). For this reason, a third module in the DER is designed to represent mood states.

Mood can be represented on one dimension good/bad as described by Reilly (1996), Picard (1997), and Kshirsagar and Magnenat-Thalmann (2002). However, in the DER described here, mood is represented by two dimensions, *calm/tense* and *energy/tiredness* — a model based on the theory of mood described by Thayer (1996). Using these two dimensions Thayer proposes four mood states: *Energetic-calm* (high energy and low tension) is the optimum mood; *Energetic-tense* (high energy and high tension) is a mood that enables people to be active and to do what has to be done; *Tired-calm* (low energy and low tension) is a relaxing state such as before sleep but a person in this state is also very sensitive to tension; and *Tired-tense*, (low energy and high tension) is the worst state when the energy is insufficient to do what has to be done.

Why could Thayer's model be more interesting than using one dimension good/bad? The main reason is that his model is a more plausible model of mood in natural systems. The dimension good/bad is generally evaluated by the difference between positive and negative emotional stimuli, and this method has two problems: in the first place, mood is "cognitively impenetrable", as argued by Sizer (2000), therefore it cannot be the result of positive or negative cognitive evaluations of events. Secondly, mood is described by Thayer (1996) and Sizer (2000) as an overview of a person's mental and physical state, not simply an overview of the emotional state. This view could also be supported by Sloman's theory of emotions if the definition of mood as a "cognitive functional architecture" given by Sizer (2000) and the definition of *meta-management mechanisms* given by Sloman (2001a) are compared. Sizer (2000) defines mood as mechanism-organising — categorising memorised information and representing the general state of a person. Sloman (2001a, p. 7) defines meta-management mechanisms as enabling "self-observation or self-monitoring of a wide variety of internal states, along with categorisation and evaluation of those states, linked to high level mechanisms for learning and for controlling future processes."

Thayer's model represents variables not normally considered emotional. The terms *energy* and *tension* are not clearly defined by Thayer (1996); they are described as a representation of the general mental and physical states of a person. The term *energy* includes mental energy and physical energy, such as the extent to which a person is awake, their blood sugar level, or their health. The energy level follows a daily biological pattern (Thayer, 1996). Tension level varies according to the pressure of the environment, such as threats, task accomplishments and maybe reproduction instinct.

By relating Thayer's model to existing work on computational emotion models, such as those described by Delgado-Mata and Aylett (2004), Cañamero (2003), and Velásquez (1997), links can be built between drives, such as hunger or thirst, and energy and tension. For instance, hunger is related to the level of energy provided by food and the need for food produces some pressure on the system to fulfil this need, which can be interpreted as tension. Thayer emphasises the existence of relationship between tension and energy. Representing mood on the basis of Thayer's model would be closer to mood theories of natural systems and could also be integrated with existing autonomous virtual agent systems.

One issue is how to reconcile Thayer's model with the suggestion of Picard (1997) to use mood state, as positive and negative, to influence the effects of emotional stimuli. Thayer's model does not represent directly a dimension good/bad or positive/negative but using his description of the four mood states two new dimensions appear:

- Pleasure/displeasure represented by the extremes: *energy-calm* (high energy and low tension), and *tired-tense* (low energy and high tension).
- Sleep/arousal represented by the extremes: *tired-calm* (low energy and low tension), and *energy-tense* (high energy and high tension).

These two dimensions are used by Russell (1997) to describe emotional states.

A positive/negative value of mood is computed as the nearest distance between the pair (tension,energy) and the line where tension=energy. If the tension exceeds the energy the mood is negative. If the tension is less than the energy the mood is positive. The computed distance represents how positive or negative the mood is and it is used in the DER to influence dynamic filters. The influences of mood enable the DER to react differently to emotional impulses, and enables the EE-FAS to produce different facial expressions according to the long term emotion history. The influence of mood is shown in Figures 4-5 and 5-1 and explained further in the next chapter.

4.3.2 The DER in a Virtual Actor Architecture

Even if the DER described previously has been designed to be integrated in the EE-FAS, it could also be introduced in a more complex virtual actor architecture. Its foundations are based on emotion theories which can be applied to other characteristics of virtual actors than facial expressions. The DER described here can be used as a mechanism influencing cognitive processes, memory recalls, action selection, perception and so on. The types of state represented by the DER can influence these mechanisms with different degree of emergency. Behaviour activations represent states that need to be taken into consideration immediately. Emotions provide contexts in which current matters should be cognitively processed. Finally, moods influence any matters by providing a bias to memory recalls, perceptions and cognitive processes. The DER could be connected to different mechanisms eliciting emotions. For instance, this DER could be integrated in an architecture such as the one described by André et al. (1999). In this architecture, the reactive and the deliberative channel could produce emotional impulses triggering behaviour activations and producing emotional states, respectively.

4.3.3 Influences of Emotional Impulses on the DER

This example shows that the state of the DER is modified by emotional impulses but these modifications depend on its original state. Figure 4-5 shows four graphs, where the bottom one represents emotional impulses sent to the EE-FAS DER and the three others represent the responses of the DER to these emotional impulses within three mood contexts; from top to bottom: negative mood marked *a*, neutral mood marked *b*, and positive mood marked *c*. The graphs *a*, *b* and *c* show the responses of the behaviour activations *Angry* and *Happy*; the responses of the dimension *happiness* and *anger*, and the intensities of the two dimensions *tension* and *energy* representing moods.

On the graph representing the emotional impulses, Figure 4-5 (d), three *happiness* impulses with different intensities arrive at 2 second intervals. Following that, three *anger* impulses arrive at 2 second intervals. Emotional impulses of *happiness* can trigger the *Happy* behaviour, increase the level of *happiness* and reduce the level of *anger*. Emotional impulses of *anger* can trigger the *Angry* behaviour, increase the level of *anger* and reduce the level of *happiness*. According to the state of the DER, the effects of the same series of emotional impulses are different.

Mood state modifies the intensity of emotional impulses in a non-linear fashion. Mood increases the low intensities of emotional impulse having the same valence and decreases all intensities of emotional impulse having opposite valence. For instance in a positive mood, the intensity of a *happiness* impulse (positive valence), is increased, whereas the intensity of an *anger* impulse is decreased because it has negative valence. These modifications are reflected in the intensities of behaviour activations, if they are triggered. In general, the three graphs *a*, *b* and *c* show that emotional impulses having the same valence as the mood valence

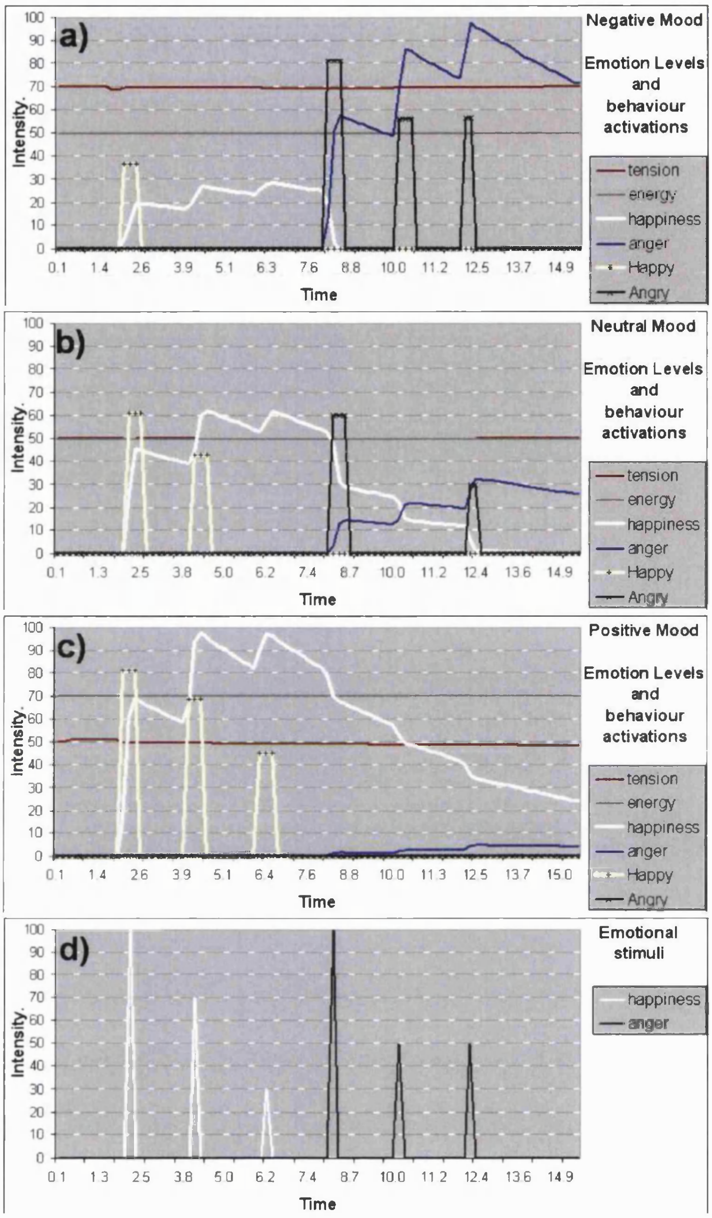

Figure 4-5: Changes of DER states due to six emotional impulses in three contexts. Graphs *a*, *b* and *c* show state changes in the contexts of *Negative Mood*, *Neutral Mood* and *Positive Mood*, respectively. Graph *d* shows the emotional impulses sent over time to the DER, where graphs *a*, *b*, and *c* show the responses of the behaviour activations *Happy* and *Angry*, the emotions *happiness* and *anger* and the dimension representing mood state.

trigger behaviours more easily than those having opposite valence. This is due to the fact that when the mood increases an impulse intensity, there are more chances for this intensity to reach the activation threshold of the corresponding behaviour.

In the case of positive mood, graph *c*, each *happiness* impulse triggers a behaviour because their intensities have generally been increased. In the case of neutral mood, graph *b*, the mood has less effect on the emotional impulse intensities. Therefore the third *happiness* impulse does not trigger the behaviour because its intensity is lower than those of the other *happiness* impulses. In the case of negative mood, graph *a*, only the *happiness* impulses having the higher intensity, the first one, triggers a behaviour because other intensities are under the activation threshold due to the reductive effect of mood. It should be noticed that the intensities of the emotions *happiness* and *anger* are also influenced differently due to the effects of mood on the intensity of emotional impulses. The three graphs *a*, *b* and *c* show that *anger* emotional impulses reduce the intensity of the emotion *happiness*.

The behaviour activation thresholds are variable. They are influenced by the intensity of emotions represented in the second module of the DER. For instance, the intensity of the emotion *happiness* increases the activation threshold of the behaviour *Angry*. This is one of the reasons why no *Angry* behaviour is activated in the context of positive mood (Figure 4-5(c)). This effect is more observable in graph *b*, neutral mood. The second and third *anger* impulses have the same intensity and the mood state is constant between these two stimuli. The second *anger* impulse does not trigger a behaviour because the threshold is increased due the intensity of the emotion *happiness*. The third *anger* impulse triggers a behaviour because the activation threshold has been decreased according to the decreased intensity of the emotion *happiness* due to the second *anger* impulse.

Anger impulses increase the intensity of the emotion *anger* and decrease the intensity of the emotion *happiness*. These effects are not constant: they are related to the intensity of the emotions. For instance, with care it is possible to see in Figure 4-5 (b) that the third *anger* impulse increases the intensity of the emotion *anger* more than the second *anger* impulse, and this is due to the lower level of the emotion *happiness* when the third *anger* impulse arrives. For the same reason, the effect of the third *angry* impulse on the emotion *happiness* is more important than the effect of the second one.

This example shows that the DER represents persisting states that can be used to influence the production of emotional facial expressions.

4.3.4 XML Configuration of the DER

The DER model can be composed of any number of modules, each one representing an emotional space. Each module can be composed of any number of dimensions. Figure 4-6 shows the XML configuration of the DER composed of three modules: mood, emotions and behaviour activations. This figure also shows that each module is composed of a list of dimensions.

```
<DERModel>
  <module name="mood">?
  + <dimension name="tension" >
  + <dimension name="energy" >
  </module>
  <module name="emotions">
  + <dimension name="happiness" >
  + <dimension name="anger" >
  + <dimension name="sadness" >
  + <dimension name="surprise" >
  + <dimension name="fear" >
  + <dimension name="disgust" >
  </module>
  <module name="behaviour activations">
  + <emotional_perception>
  + <dimension name="Happy" >
  + <dimension name="Angry" >
  + <dimension name="Sad" >
  + <dimension name="Surprised" >
  + <dimension name="Afraid" >
  + <dimension name="Disgusted" >
  </module>
</DERModel>
```

Figure 4-6: XML fragment of the configuration for the DER containing three modules

Figure 4-7 shows the XML fragment of the configuration for the dimension happiness. In this figure, four dynamic filters are shown, each one is used to control the effects of a particular type of emotional stimuli. In the order in which the dynamic filters appear in the figure, they control the effects of happiness, anger, sadness and fear stimuli. Within each dynamic filter, a node `<influence>` defines the influences of other dimensions on this dynamic filter.

```
- <dimension name="happiness" >
   - <filters>
     - <dynamic_filter>
       + <function name="Happ. Stim." type="sigmoid"
               input_type="stimulus" selective_para="name"
               selective_value="happiness" input_para="peak intensity">
       + <influence>
     </dynamic_filter>
     - <dynamic_filter>
       + <function name="Anger Stim." type="sigmoid"
               input_type="stimulus" selective_para="name"
               selective_value="anger" input_para="peak intensity">
       + <influence>
     </dynamic_filter>
     - <dynamic_filter>
       + <function name="Sadn Stim." type="sigmoid"
               input_type="stimulus" selective_para="name"
               selective_value="sadness" input_para="peak intensity">
       + <influence>
     </dynamic_filter>
     - <dynamic_filter>
       + <function name="Fear Stim."  type="sigmoid"
               filter_what="stimulus" selective_para="name"
               selective_value="fear" input_para="peak intensity">
       + <influence>
     </dynamic_filter>
     ...
  </filters>
</dimension>
```

Figure 4-7: XML fragment of the configuration of the dimension happiness.

The DER model defines *dynamic filters* as functions which have parameters varying over time and which are used to modify values of emotional stimulus parameters. A *dynamic filter* is composed of a *function* with its list of parameters and a series of *influences* that modify the *function* parameters. The aim of the function is to take the value of one emotional stimulus parameter as an input and to provide a new value. The definition of influences also requires a function that takes the intensity of a dimension as an input and provides a corresponding value that will be used to modify a parameter of the dynamic filter function. To design a DER model that can be configurable by users a *generic function* type is defined with the following parameters:

- *name:* is a name describing the function.
- *type:* is the type of mathematical function used to define the function: linear or sigmoid.

```
- <dynamic_filter>
  - <function name="Happ. Stim." type="sigmoid"
            input_type="stimulus" selective_para="name"
            selective_value="happiness" input_para="peak intensity">
      <gain>100</gain>
      <down_up_shift>0</down_up_shift>
      <right_left_shift>60</right_left_shift>
      <slope_steepness>20</slope_steepness>
    </function>
  - <influence>
      <influenced_parameter>right_left_shift</influenced_parameter>
      <operation>addition</operation>
      <inflencing_factor>1</inflencing_factor>
    - <influencing_source type="dimension" operation="abosulte maximum">
      + <function name="anger influence" type="sigmoid"
                  input_type="dimension" selective_para="name"
                  selective_value="anger" input_para="intensity">
      + <function name="Sadn. influence" type="sigmoid"
                  input_type="dimension" selective_para="name"
                  selective_value="sadness" input_para="intensity">
      + <function name="fear influence" type="sigmoid"
                  input_type="dimension" selective_para="name"
                  selective_value="fear" input_para="intensity">
      </influencing_source>
    </influence>
</dynamic_filter>
```

Figure 4-8: XML configuration of a dynamic filter.

- *input_type:* gives the type of object taken as input by the function: emotional stimulus or dimension. This is the type of object that is modified by this function.

- *selective_para:* indicates the name of an object's characteristic that is used to *select* which object should be modified. The characteristics for an emotional stimulus could be *name*, *valence*, *decay duration* and so on; for a dimension it is its *name*.

- *selective_value:* indicates the value of the characteristic specified by *selective_para* used to *select* the object that should be taken as input.

- *input_para:* indicates the name of the object's characteristic that should be modified by the function, such as *peak intensity* or *sustain duration.*

Depending on the mathematical type of the function, a list of parameters should be defined. For instance, in the case of a sigmoid function four parameters are defined: *gain*, *down/up shift*, *right/left shift*, and *slope steepness.*

In the case of the function of a dynamic filter, the function's parameters can be modified by *influences*. An *influence* modifies one parameter of the dynamic filter function and is defined with the following parameters:

- *influenced_parameter:* indicating which parameter of the dynamic filter function should be modified, such as *right/left shift* for a sigmoid function.
- *operation:* specifies which operation should be used to modify the function parameter. In the current implementation two operations are possible *addition* and *replace.*
- *influencing_factor:* indicating by which factor the influencing source should modify the function parameter.

As well as these parameters, a list of *influencing sources* and an operation to combine the values of the *influencing sources* need to be defined. An *influencing source* is defined by *generic functions* as described previously. Figure 4-8 shows the configuration of a dynamic filter controlling the effect of happiness stimuli by modifying its peak intensity according to the intensity of the emotions anger, sadness and fear.

4.4 Conclusion

This Chapter focuses on Dynamic Emotion Representations, in contrast to mechanisms eliciting emotions. Dynamic Emotion Representations maintain states changing over time and are influenced by mechanisms eliciting emotions through emotional impulses. Dynamic Emotion Representations are general mechanisms modifying the results of emotional assessment independently of the causal events.

The Dynamic Emotion Representation model developed for this thesis is inspired by Picard's description of emotional stimuli, emotion intensity and dynamic filters. Using the DER model it is possible to represent any number of emotional spaces, such as moods, emotions, or even drives. Each emotional space is composed of any number of dimensions. In the DER model, any emotional impulses can affect any dimensions and the effects of emotional impulses are controlled through dynamic filters. These dynamic filters are influenced by the state of the DER, binding the effects of emotional impulses to this state. The creation of an instance of the DER model is carried out using an XML file, as described in this Chapter.

An instance of the DER model has been built to represent three types of states changing on different timescales. The choice of the states represented by this DER is based on emotion theories, as well as on the needs of the EE-FAS for the production of facial expressions. Even though the DER has been designed for the EE-FAS, it could be used in a larger virtual actor architecture.

It has been shown that the DER states are persistent, representing emotional momentums. According to the original state of the DER, emotional impulses affect its states differently.

The instance of the DER model described in this chapter is integrated into a Emotionally Expressive Facial Animation System to produce and select facial expressions. The Emotionally Expressive Facial Animation System and the integration of the DER presented in this Chapter, is described in the following Chapter.

Chapter 5

The Emotionally Expressive Facial Animation System

5.1 Introduction

This research aims to design and implement an animation system enabling non-professional animators or other programs to create facial animations using the meanings of facial expressions.

Four Levels of Control (LoCs) for creating facial animations have been defined in Chapter 1. The higher the LoC, the lower the expert knowledge animators need for the production of visual speech. Here is a summary of the four Levels of Control:

1. *Static Graphical Representation*, such as 2D, 3D mesh or volumetric representations.
2. *Graphical Deformation Commands or Animation Parameters*, such as Bézier patches, Abstract Muscle models, Facial Animation Markup Languages and the Facial Action Parameters defined by MPEG-4 (Parke and Waters, 1996; Hartung et al., 1998; Pirker and Krenn, 2002; Gustavsson et al., 2001).
3. *Facial Movements or Facial Signals* are based on the same descriptions as

the previous LoCs but, in addition, functions of time are used to describe the changes of parameter intensity over time.

4. *Facial Meanings* are not physical descriptions of the face but instead they describe meanings expressed through the face.

Level of Control 4 is the most intuitive LoC because it does not require any knowledge about graphical or physical configurations of the face. Instead, it enables animators to describe the face with meanings that need to be communicated. For instance, it is more intuitive to use a command saying "emphasise this word" than to describe the facial expression that should be used for this function. LoC4 is the most appropriate of the four LoCs defined to build a facial animation system that can be used by non-specialist users or other programs.

One issue arising from the use of facial meanings at a user level is the transfer of the complexity from the user into the system. The transformation from facial meanings into visual representation is not direct: "The human face can generated around 50,000 distinct facial expressions, which correspond to about 30 semantic distinctions" (Paradiso, 2002, page 1).

How do we map a small number of categories of facial meanings onto a large number of facial expressions?

This Chapter presents the Emotionally Expressive Facial Animation system (EE-FAS). The EE-FAS integrates in its architecture the DER described in Chapter 4 to extend the number of facial expressions that it is able to display.
The EE-FAS:

- distinguishes and produces two types of facial expression
  - When the character is not speaking, the six classical Ekman expressions are used to show the effect of an emotional event on the character. Universally recognised facial expressions are generally considered as full-face configurations, therefore they cannot be combine with speech related expressions.
  - When the character is speaking, the visual representation of the face is affected by scripted communicative acts (e.g. to emphasise a word).

Facial expressions corresponding to communicative acts are synchronised with and support the speech. They are considered as partial face configurations that can be combined with other speech related facial expressions.

- produces different facial expressions corresponding to one communicative acts according to the emotional state of the virtual actor.
- and produces masking facial expressions and facial expressions corresponding to blend emotions.

Based on these concepts the EE-FAS:

- provides users with functions of communicative acts and emotional impulses to produce facial animations;
- produces different animations from the same script, according to the emotional history of the virtual actor;
- is not only scripted but also reacts to unscripted emotional events coming from the environment through emotional impulses.

The EE-FAS has also the advantage of using an extensible, modular blackboard-based architecture and to defined its facial expressions in customisable XML files.

This Chapter is composed of Section 5.2 describing the use of facial meanings in facial animation systems; Section 5.3 presenting some animations produced by the EE-FAS from animation scripts played within different emotional contexts; Section 5.4 presenting the architecture of the EE-FAS and Section 5.5 explaining how facial signals and visemes are synchronised with the speech.

5.2 Facial Meanings and Animation Systems

This section describes the different categorisations of facial meanings and how they are used in existing facial animation systems. It also explains the facial meaning categorisation built for the EE-FAS.

5.2.1 Facial Meanings and Facial Animation Systems

Several categorisations of facial meanings exist. Ekman (1979) distinguishes: *emblems* replacing the speech, *emotional emblems* showing non-felt emotions, *conversational signals* like emphasisers, *punctuators* marking pauses in the speech, *regulators* used to organise the dialogue, *manipulators* responding to biological needs such as blinking, and *affective displays* which are the displays of felt emotions. Pelachaud et al. (1991), Albrecht et al. (2005), The Duy Bui (2004), and Thórisson (1999) describe facial animation systems based on this description.

Some facial animation systems focus on the display of emotional expressions based on a set of universally recognised expressions (Ekman, 1992; Izard, 1971). Certain systems display pure versions of these facial expressions (Velásquez, 1997; Kshirsagar and Magnenat-Thalmann, 2002; Paiva et al., 2004) and other extend the number of facial expressions by combining them (Kurlander et al., 1996; Latta et al., 2002; Raouzaiou et al., 2003; The Duy Bui, 2004; Albrecht et al., 2005).

Pelachaud and Bilvi (2003) categorise facial expressions as communicative functions according to *belief*, *intention*, *affective state* and *metacognitive*.

It has been suggested that a communicative function can have different visual representations, i.e. facial signals, according to the physical and mental states of the speaker, and to the state of the dialogue (Bavelas and Chovil, 2000; Pelachaud and Bilvi, 2003). Cassell et al. (2001) describe a system generating multiple facial signals corresponding to syntactic or contextual information related to the text to be spoken. These facial signals are filtered according to eventual conflicts between them and their priorities. The system called Ymir (Thórisson, 1999) can produce multiple facial signals for one communicative function. Instead of using the physical state of the virtual actor to select a signal, like Ymir, the EE-FAS uses its emotional state. The EE-FAS is also different from a system like FacEMOTE (Badler et al., 2002), which modifies an already selected facial signal, because the EE-FAS selects from various facial signals according to the emotional state.

5.2.2 Emotional Expressions and Emotion Representations

Existing facial animation systems use *static* (Albrecht et al., 2005; Kurlander et al., 1996; Raouzaiou et al., 2003; Latta et al., 2002) or *dynamic* (Kshirsagar and Magnenat-Thalmann, 2002; The Duy Bui, 2004; Velásquez, 1997; Paiva et al., 2004; Reilly, 1996) emotion representations to select which emotional expression should be displayed.

Static emotion representations, such as the *emotional wheel* described by Plutchik (1980), are used to map emotion words to combinations of universally recognised facial expressions (Albrecht et al., 2005; Kurlander et al., 1996; Raouzaiou et al., 2003; Latta et al., 2002). The emotional wheel describes an emotional space with two dimensions: activation and evaluation. In this space are placed basic emotions corresponding to the universally recognised facial expressions. When a new emotion needs to be expressed, it is placed on the emotional space using the two dimensions and the universally recognised facial expressions of the two closest basic emotions are combined to create the expression of this new emotion.

One issue with static emotion representations is the lack of a consistency mechanism. Emotions vary relatively slowly, so a change of expression from happiness to anger takes time. Static emotion representations do not provide any consistency mechanism so any emotional expression can be displayed at any time, regardless of the previous emotional expressions.

Dynamic Emotion Representations keep track over time of changes in emotional states and thus represent emotional momentums. Emotional expressions due to the changes of state of a Dynamic Emotion Representation are consistent between each other over time and show believable behaviour.

5.2.3 Facial Meanings in the EE-FAS

The EE-FAS is based on Ekman's categorisation of facial expressions (Ekman, 1979) and a description of communicative acts given by Bavelas and Chovil (1997). A summary of the facial meanings used by the EE-FAS is shown Table 5.1. The EE-FAS distinguishes *manipulators* from other facial meanings be-

Table 5.1: Comparison between the categorisations of facial meanings used by the EE-FAS and described by Ekman (1979).

EE-FAS	Ekman
Manipulators	Manipulators
Emotional expressions	Affective displays
Communicative acts	
Portrayal	Emblems, emotional emblems
Thinking	
Personal reaction	
Felt emotions	Affective displays
Not felt emotions	Emotional emblems
Grammatical markers	Communicative signals, punctuators
Question marker	
Emphasiser	
Organisational of story	Regulators

cause they are reactions to body needs. *Emotional expressions* are defined here as involuntary, but more or less controllable, reactions to emotional events. In this thesis, emotional expressions are the universally recognised expressions and they are defined as occurring at the *emotion activation* (Russel and Fernández-Dols, 1997), such as a facial expression of surprise, or anger. In the EE-FAS, emotional expressions are displayed as reactions to emotional impulses; they are the classical Ekman expressions. These emotional expressions are not related to the speech, they can even interrupt it. They are used in the EE-FAS to communicate a clear single emotional message as a fast reaction to a change in the environment. For example, the virtual actor can be interrupted during its speech and display a facial expression of surprise. Universally recognised facial expressions cannot be displayed during the speech due to their visual/physical definitions: the expression of anger involves tight lips, the expressions of surprise and fear involve an open mouth, and the movement of lip corners in the expressions of happiness and sadness conflict with certain visemes. *Communicative acts* are facial expressions related to and synchronised with the speech. Communicative acts have two components, a communicative function, such as emphasising words, and a signal, which is the visible movement of the face, such as raising eyebrows (Bavelas and Chovil, 2000; Pelachaud and Bilvi, 2003). Communicative acts are symbolic and acted, in contrast to those due to emotional reactions (Chovil, 1997).

Emotional expressions communicate their emotional meanings through full-face configurations, where communicative acts can express meanings through partial-face configurations (Chovil, 1997; Smith and Scott, 1997). The EE-FAS uses the definition of communicative functions given by Bavelas and Chovil (1997). This definition is closer to the structure of the sentence than the description given by Pelachaud and Bilvi (2003), therefore it should be easier for users to mark-up the text to be spoken. The EE-FAS also distinguishes two types of *personal reaction*, the first type is when a person does not feel the emotion that he tries to express, and the second one is when the person feels the expressed emotion. Facial signals corresponding to personal reactions when the person does not experience the emotion are different from those corresponding to *portrayal* because they try to be believable where portrayal could be exaggerated and symbolic.

An example illustrating the differences between an emotional expression and a personal reaction could be the differences between a smile due to a joke and a polite smile to acknowledge a person. In the first case, the smile can interrupt the speech, it is not directly controlled by the person, and it is due to an emotional event. In the second case it can be displayed during the speech and it is a deliberate action. Depending on the emotional state of the person, the second type of smile can be "genuine" or "fake" (Ekman, 1992).

5.2.4 Production of Fake and Genuine Expressions

The theme of genuine and fake expressions is rarely touched in computer animations. This is probably due to the difficulties of displaying subtle differences between genuine and fake expressions. For instance, the difference between a fake smile and a genuine smile is situated around the eyes; the eyes should smile with the rest of the face (Ekman, 1992). The movements around the eyes are due to muscles only activated during a happy emotional episode.

The Duy Bui (2004) suggests having two emotional states; one is the emotional state felt by the virtual actor and the other represents the emotional state that should be displayed. If these two states are different, it can be said that the agent fakes its emotional expressions. In this case the agent displays the same facial expressions when it hides its felt emotions as when it shows its felt emotions.

This virtual actor is a perfect liar, not showing any differences between genuine and fake emotional expressions.

In this thesis, a fake expression is an facial expression displayed to mask the display of a felt emotions. The masking expression is given by the communicative function as a deliberative action. A genuine expression is displayed when the deliberative expression represented by a communicative act matches the emotional state of the virtual actor. In the EE-FAS, an intended emotional meaning can be specified through a particular type of communicative function called a *personal reaction* and the emotional state is given by the DER. By comparing the state of the DER and the emotional value of communicative functions, the EE-FAS can distinguish contexts in which a genuine or a fake facial expression should be displayed. An example of fake and genuine facial expressions is shown by the screen-shots of the third row in Figure 5-4 and will be discussed in Section 5.3.2.

5.3 Emotionally Driven Animations

This section presents two examples of facial animations produced by two different mechanisms implemented in the EE-FAS. The first example is series of emotional facial expressions due to successive emotional impulses in different mood settings. The second example is based on an animation script played in different states of the DER. Both examples show the influences of the DER state on the production of facial expressions.

5.3.1 Emotional Expressions in Emotional Contexts

This example shows the production of Ekman's emotional facial expressions due to emotional impulses. According to the state of the DER, emotional impulses trigger behaviour activations with a certain intensity which are reflected in the strength of the displayed facial expressions. However the intensity of the behaviour activation, and in consequence the intensity of the facial expressions, is influenced by the mood state. In Figure 5-1, rows correspond to different mood settings: top row is negative mood, middle row is neutral mood and bottom row

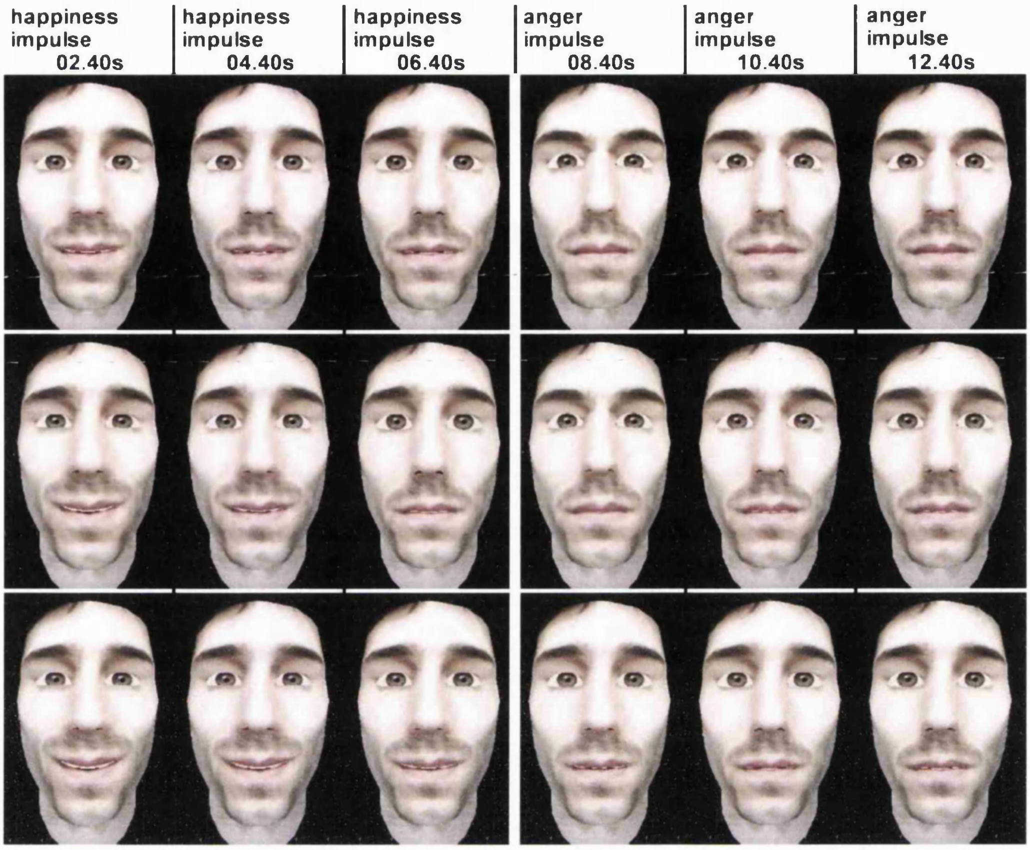

Figure 5-1: First, second and third row show screen-shots in the contexts of *Negative Mood*, *Neutral Mood* and *Positive Mood*, respectively. Each column contains screen-shots at the time when an emotional impulse is sent to the DER.

is positive mood. In this Figure, columns are screen-shots taken when an emotional impulse is sent. Through time, six emotional impulses are sent at 2 second intervals; first three are happiness impulses and the next three are anger impulses. This example is based on the same settings as explained in Section 4.3.3 describing the responses of the DER. Here the focus is on the visual responses produced by the EE-FAS. However, Figure 5-2 is repeated in this section as a reference to emphasise the differences between the background behaviour of the DER and the visual behaviour of the EE-FAS.

The happiness emotional impulses, having the intensities of 100%, 70% and 30%, are applied in three mood contexts.

In row 1, the first screen-shot shows some response to the happiness impulse. However this response is reduced due to the influence of negative mood. The two other screen-shots do not shows any response because the impulse intensities are to low.

In the context of neutral mood, row 2, the first happiness impulse produces a stronger response than in the context of negative mood. In addition the second happiness impulse also produces a response but the intensity of the third impulse is still too low to generate any response.

The positive mood, row 3, has an amplification effect on happiness impulses, therefore they all produce facial expressions, and these expressions have stronger intensities than in the previous mood contexts.

The effects of anger impulses are influenced by mood states but in addition, they are also influenced by the level of the emotion happiness generated by the previous happiness impulses.

In row 1, the three anger impulses produce responses, showing an expression of anger represented by an eyebrow frown and tight lips. The different impulse intensities, 100%, 50% and 50%, produce different strength of expressions but the differences are not marked as strongly as excepted due to the effect of negative mood amplifying more lower intensity impulses.

In row 2, representing the context of neutral mood, shows responses when mood has less influence on anger impulses than in the context of negative mood. The response of the first anger impulse is more important than the responses of the two other anger impulses, which both have of the same intensity. In fact, Figure 5-2 shows that no behaviour activation has been triggered by the second anger impulse, therefore no facial expression should be displayed. The expression shown at the time of the second anger impulse is due to the visual momentum of the expression produced by the first anger impulse. The third anger impulse does produce an emotional expression. The only difference influencing the effects of the second and third anger impulses is the emotional momentum produced by the previous happiness impulses. Happiness impulses increase the level of happiness, which takes time to disappear and reduces the effects of anger impulses.

In the context of positive mood, row 3, anger impulses do not produce any response. This is due to two reasons. In the first place, the positive mood reduces the effects of anger impulses. Secondly, the happiness momentum produced by the previous happiness impulses also reduce the effects of anger impulses.

This example, supported by Figure 5-1, shows that emotional expressions are produced in a coherent fashion due to the momentums of the DER. The display of emotional expressions is dependent on the state of the DER, i.e. on the emotional history.

5.3.2 Communicative Functions in Emotional Contexts

The communicative function tags are used to produce facial movements synchronised with the speech. In contrast with existing systems, the EE-FAS use the emotional state of the virtual actor to select facial signals corresponding to communicative functions.

By changing the emotional state of the virtual actor, an animation script, shown in Figure 5-3, produces different animations. Figure 5-4 shows screen-shots of two animations resulting from the same animation script but with different states of the DER. The four pictures in the first row are results in the context *happiness* whereas the four pictures in the second row are results in the context *sadness*.

Each screen-shot within a row in Figure 5-4 corresponds to one of the four tags described below.

Communicative Function Expressing Emotions: The following tag permits a facial expression showing the emotional state of the virtual actor. The results of this tag are shown by the two screen-shots at 01.30s in Figure 5-4: top picture when the virtual actor is happy and the bottom one when it is sad.

```
<comm_func name="personal reaction" intensity="60" >
```

Communicative Function Emphasiser: The following tag is used to emphasise a word or series of words.

```
<comm_func name="Emphasizer" intensity="30">
```

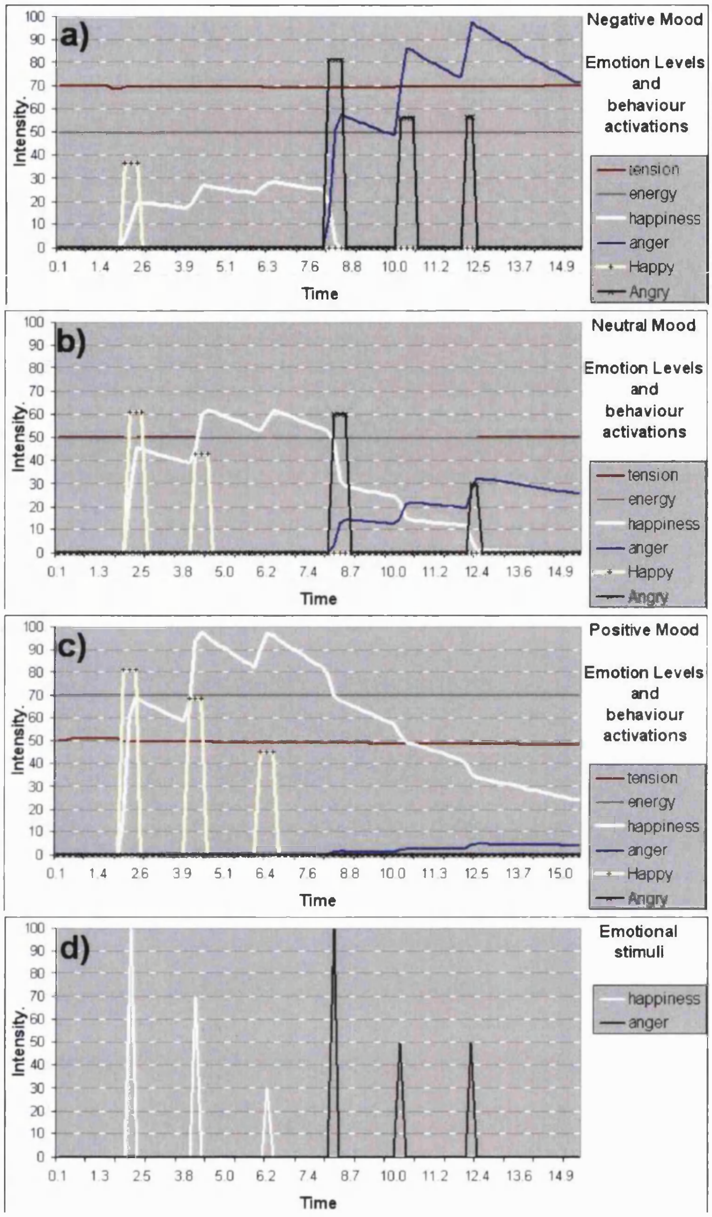

Figure 5-2: Changes of DER states due to six emotional impulses in three contexts. Graphs *a*, *b* and *c* show state changes in the contexts of *Negative Mood*, *Neutral Mood* and *Positive Mood*, respectively. Graph *d* shows the emotional impulses sent over time to the DER, where graphs *a*, *b*, and *c* show the responses of the behaviour activations *Happy* and *Angry*, the emotions *happiness* and *anger* and the dimension representing mood state.

```
<input>
  <text>
    <comm_func name="personal reaction" intensity="60">
      You can blame the Sirius Cybernetics Corporation for making androids
      <comm_func name="Emphasizer" intensity="30">
         with GPP...
      </comm_func>
    </comm_func>
  </text>
  <silence duration="1" />
  <text>
    <comm_func name="personal reaction" max_nb_fd="2" sec_emo_context="happiness" intensity="60">
      Genuine People Personalities.
    </comm_func>
  </text>
  <silence duration="0.4" />
  <text>
    <comm_func name="personal reaction" intensity="60">
      I'm a personality prototype. You can tell,
      <comm_func name="question marker" intensity="60">
        can't you...?
      </comm_func>
    </comm_func>
  </text>
</input>
```

Figure 5-3: Animation script used to produce the animations related to the screen-shots in Figure 5-4.

The facial signals used for this communicative function are selected according to the emotional state of the virtual actor. In the screen-shots, the facial signals corresponding to the communicative function are combined with the previous communicative function *personal reaction*. In the context of *happiness*, an *eyebrow raise* is the facial signal used to emphasise words, whereas in the context *sadness* an *oblique movement of the eyebrows* is used for this.

In this example, the facial signal is not as visible in the context of *happiness* as in the context of *sadness*. This is due to the fact that these facial signals are additive. In the context of *sadness*, the same facial signal, e.g. *eyebrow oblique*, is used by both the previous communicative function and the current one. One advantage of selecting facial signals according to the emotional state is the reduction of eventual conflicts between signals; in this example the same movement of facial part is used, eliminating the possibility of conflicts.

Masking and blending v.s. genuine expression: The following tag produces the expressions shown by the screen-shots at 10.00s, Figure 5-4.

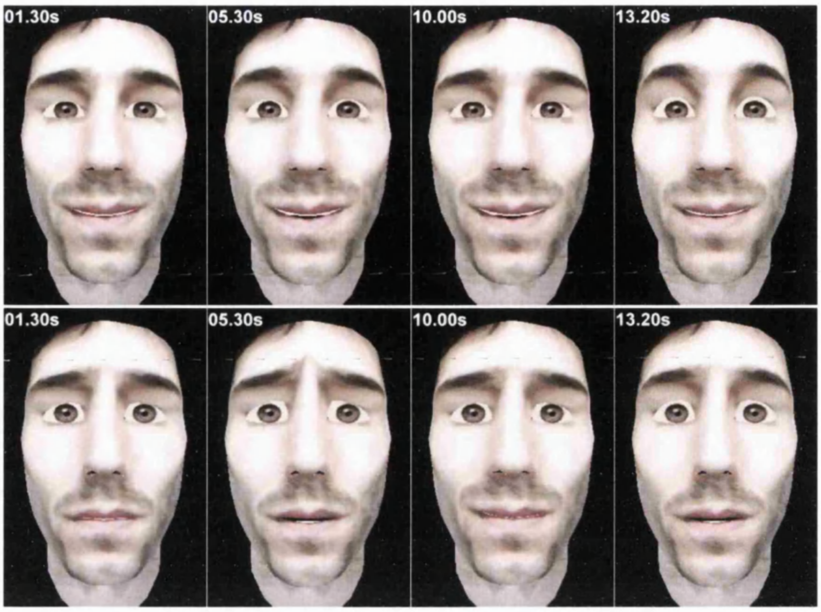

Figure 5-4: Four screen-shots of two animations resulting from the same animation script played in two different emotional contexts. The top four pictures are in a context *happiness* whereas the four bottom pictures are in a context *sadness*.

```
<comm_func name="personal reaction" max_nb_fd="2"
 sec_emo_context="happiness" intensity="60" >
```

Due to the attribute `sec_emo_context="happiness"`, the virtual actor displays a smile in both contexts. However, in the context of *happiness* the smile is "genuine" whereas in the context of *sadness* it is "fake". The genuine smile is represented by lip corners raised, lower eyelids raised, and an open mouth. The fake smile is represented by an asymmetric raising of the lip corners, the masking expression, combined with the expression of the "felt" emotion, in this case the expression of sadness.

The masking expression is given by the attribute `sec_emo_context` and the felt emotion give by the DER is displayed due to the attribute `max_nb_fd`. The latter attribute specifies whether one or two facial signals should be displayed at the same time.

Communicative Function Question Marker: The final screen-shots at 13.20s in Figure 5-4 are the products of the communicative function *question marker.* In the context *happiness*, the facial signal used is *eyebrows raised.* In the context *sadness*, it is a combination of *eyebrows raised* and *eyebrows oblique.* This combination is not due to the presence of a previously displayed facial signal, as in the screen-shots at 5.30s, but it is due to the signals chosen by the user to express a *question marker* in the context of sadness.

```
<comm_func name="question marker" intensity="60" >
```

It should be emphasised that the definition of the facial signals corresponding to communicative function can be modified by altering an XML file. The facial signals chosen to fulfil communicative functions are not based on any particular model but they use to prove a concept.

These examples show that different facial animations can be created from the same script according to the emotional state of the virtual actor.

5.4 The Architecture of EE-FAS

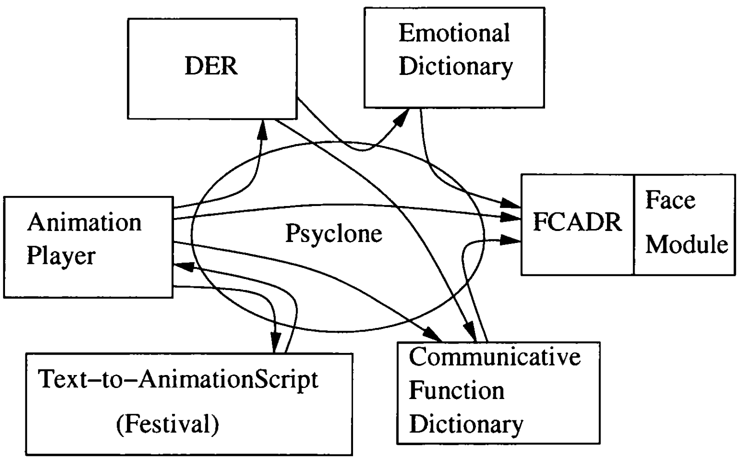

Figure 5-5: Emotionally Expressive Facial Animation System Architecture.

5.4.1 Transformations of Meanings onto the Screen

In the EE-FAS, facial meanings can be emotions communicated by emotional expressions, or functions of communicative acts. Before being displayed on the screen, facial meanings go through several transformations and the results of each transformation can be seen as commands at a different Level of Control.

The advantage of using several Levels of Control within a system is the possibility to split the complexity of commands into several steps and to design a flexible system. Instead of directly transforming a facial meaning into positions of points on the screen, the use of Levels of Control enables the system to transform a facial meaning into a description of movements of facial parts, then a movement of a facial part into contractions of Abstract Muscles, and a contraction of an Abstract Muscle into positions of vertices in a 3D space. This implementation also makes the system more flexible because each Level of Control is independent of the others. Therefore, if a different description is used for one Level of Control, all the commands at higher Levels of Control are still valid, only the transformation to the new description needs to be redesigned. For instance, an animation using meanings of facial expression can be applied to a dog as well as a human actor, only the physical descriptions need to be changed.

An interesting characteristic of using different Levels of Control is the loose transformations between one level to another. Communicative functions are transformed into facial signals but this transformation is dependant on the emotional state of the virtual actor. The loose transformations from one Level of Control to another are also used to smooth the movements between two facial signals by limiting the speed at which Abstract Muscle contractions can change values.

In the EE-FAS, each transformation from one Level of Control to another is carried out by a module based on a customisable XML dictionary.

5.4.2 Message Flow in the EE-FAS

The EE-FAS has been implemented with a modular architecture where modules communicate by sending messages to blackboards, through middleware called

Psyclone (Thórisson, 2004; Thórisson et al., 2004). To help explain how the EE-FAS works, the flow of messages occurring for the production of an animation is described here. Figure 5-5 shows the architecture of the EE-FAS, including all the modules and the work flow of the messages used for the communication between modules.

The input module, called the *Animation Player*, takes a tagged text as input. Tags are introduced within the text by the animator and can describe *communicative functions* and *emotional impulses.* In this particular example only communicative function tags are used.

```
<comm_func name="personal reaction" intensity="60">
 You can blame Sirius Cybernetics Corporation for making androids
 <comm_func name="Emphasizer" intensity="30" >
 with GPP...
 </comm_func>
</comm_func>
```

This tagged text is sent to the module called *Text-To-AnimationScript* which transforms it into an animation script. An animation script is composed of the list of tags, including viseme tags (visual counterparts of phonemes), communicative function tags and emotional impulse tags. All these tags have time information specifying when they should be displayed and synchronising them with the speech. The timing information is given by the Text-To-Speech software called Festival, which provides the timing of the phonemes (CSTR, 2006).

The animation script is sent to the *Animation Player* module which can now produce the animation. Each type of tag is sent to different modules; viseme tags are directly sent to the module called *Facial Component Action Dynamic Representation* (FCADR); communicative function and emotional impulse tags are sent respectively to the modules called *DER* and *Communicative Function Dictionary.*

The *FCADR* module is discussed in more detail in Appendix B; its function is to keep track of the current facial configuration, to combine facial signals coming from other modules, and to send facial descriptions to the *Face* module, which projects the 3D facial mesh onto the screen. A formal definition of a facial signal, operators manipulating facial signals and a set of equations to combine facial signals have been designed to cope with the combinations of facial signals and to solve eventual conflicts between them. These definitions are based on facial component actions and enable the definition of facial parts supporting one *or* multiple facial component actions at one time. The selection mechanism is based on priorities set in the definitions of facial signals. All the equations and definitions are listed in Appendix C.

5.4.3 Emotional Expressions in the EE-FAS

Emotional impulses can be introduced as tags in the text to be spoken, or they can be sent to the DER by a different application through Psyclone. According to the intensity of an emotional impulse and to the state of the DER an associated behaviour activation may be triggered. These behaviour activations are displayed as emotional expressions, e.g. one of the six universally recognised facial expressions. For instance, a strong impulse of *happiness* could produce a *Happy* behaviour activation which would be displayed as a smile. If emotional impulses are sent by another application, this mechanism would enable the virtual actor to react to changes in the environment (including to other actors) and to display a clear emotional message.

From the DER, a message representing the state of the DER is sent to a module called the *Emotional Dictionary*. If a behaviour is active, the emotional dictionary module will send an emotional expression to the FCADR (Figure 5-5). The strength of an emotional expression is directly proportional to the intensity of the triggered behaviour, which is itself related to the intensity of the emotional impulse and the state of the DER (Figures 4-5 and 5-1). The relations between behaviour activations and emotional expressions are defined in an XML file which can be customised by animators. For example, the definition of what should be displayed when the *Happy* behaviour is activated can be re-defined by the animator.

5.4.4 Communicative Functions in the EE-FAS

Communicative functions are used to produce facial movements synchronised with and supporting the speech. Communicative function tags found in the spoken text, such as those shown in Section 5.3, are sent from the *Animation Player* module to the *Communicative Function Dictionary* module. As has been discussed previously, the relation between communicative functions and facial signals is one-to-many. This mapping is carried out using the emotional state represented by the DER. For each communicative function a list of facial signals is defined and these facial signals are categorised by *emotion context* and by "fake" and "genuine" definitions. The emotional contexts are compared with the state of the DER, in particular the state of the emotion module, to select a facial signal corresponding to a communicative function. The communicative functions are defined in an XML file that can be customised by users. Figure 5-6 shows an extract from this XML dictionary.

The DER module provides emotional reports to the communicative function dictionary module. The name of the emotion with the higher intensity is used to select a facial signal corresponding to a communicative function. The animator can choose whether or not to provide an argument called *sec_emo_context* with communicative function tags. In the absence of this argument, the facial signal is selected according only to the state of the DER and to the type of communicative function. This mechanism can reduce the amount of information needed from the animators.

```
<comm_func name="personal reaction"
 sec_emo_context="happiness" intensity="60" >
```

In the example above, the virtual actor wants to display a facial expression communicating happiness but, depending on the emotional state, a genuine or a fake facial signal definition is selected. If the name of one of the two emotions with the highest intensity is equal to the value of the argument `sec_emo_context`, or if the argument is not present, a genuine facial signal definition is selected. In the case where the value of this argument does not match any emotion name with high intensity, a fake facial signal definition is selected. Section 5.3.2 gives an example of these two cases as well as other examples of communicative functions

```
<comm_function>
 <comm_func_name>personal reaction</comm_func_name>
   <contrained_by_time>no</contrained_by_time>
   - <emotional_context>                           -------------------------------------
     - <possible_emotional_states>                                              |
        <secondary_emotion>anger</secondary_emotion>                            |Def. of
     </possible_emotional_states>                                               |
     - <fake_FD_definition>                        ----------------------------- |facial
        - <FC>                                     ---------------- |            |
           <name>right_eye_brown</name>              |Facial         |            |signals
           <action_name>frown</action_name>          |Component      |Definition  |
           <attack_duration>0.05</attack_duration> |with time      |            |communicating
           ...                                       |description    |of the      |
        </FC>                                      ---------------- |            |anger
        - <FC>                                     ---------------- |the masking |
           <name>left_eye_brown</name>               |Facial         |            |as a
           <action_name>frown</action_name>          |Component      |expression  |
           <attack_duration>0.05</attack_duration> |with time      |            |communicative
           ...                                       |description    |for anger   |
        </FC>                                      ---------------- |            |act
     </fake_FD_definition>                         ----------------------------- |
     + <genuine_FD_definition>                     | Def. of genuine expression  |
   </emotional_context>                            ------------------------------------
   - <emotional_context>                           ------------------------------------
     - <possible_emotional_states>                          |
        <secondary_emotion>happiness</secondary_emotion>  |Def. of facial signals
     </possible_emotional_states>                           |communicating happiness
     + <fake_FD_definition>                                 |as a communicative act
     + <genuine_FD_definition>                              |
   </emotional_context>                            ------------------------------------
...
```

Figure 5-6: Extract of the XML file defining communicative functions using facial component actions.

Figure 5-7 shows an example of an XML message of a facial signal, corresponding to a smile, and based on facial component actions.

```
<facial_signal starting_time="" merging_mode="3">
  <facial_component combine_level="FCA" FCA_weight="2">
    <name>mouth</name>
    <action>
      <name>raise_left_lip_corner</name>
        <description>
          <peak_intensity>40.</peak_intensity>
          <delay_duration>0</delay_duration>
          <attack_duration>0.2</attack_duration>
          <sustain_duration>9.8</sustain_duration>
          <decay_duration>10</decay_duration>
        </description>
   </action>
  </facial_component>
  <facial_component combine_level="FCA" FCA_weight="2">
    <name>mouth</name>
    <action>
      <name>raise_right_lip_corner</name>
      + <description>
    </action>
  </facial_component>
  <facial_component combine_level="FC" FCA_weight="2">
    <name>right_low_eyelid</name>
    <action>
      <name>close</name>
      <description>
        <peak_intensity>28.</peak_intensity>
        <delay_duration>0</delay_duration>
        <attack_duration>0.2</attack_duration>
        <sustain_duration>9.8</sustain_duration>
        <decay_duration>10</decay_duration>
      </description>
    </action>
  </facial_component>
  ...
</facial_signal>
```

Figure 5-7: Example of an XML message describing a smile using facial component actions.

5.4.5 Facial Component Actions to Abstract Muscles

To animate the 3D facial mesh an Abstract Muscle model is used (Parke and Waters, 1996; Tanguy et al., 2003; The Duy Bui, 2004). The choice of this technique is purely due to the availability of its implementation. The animation system has been developed to control a synthetic face independently of the technique used to display the face on the screen. To gain this independence, the system manipulates facial component actions, which are closely related to Action Units

Table 5.2: Facial component, Actions, Abstract Muscles and Action Unit relations

Facial component	Action	Abstract Muscle	AU
right eyebrow	frown	right secondary frontalis	4
		right lateral corigator	
left eye brow	frown	left secondary frontalis	
		left lateral corigator	
right eyebrow	raise	right frontalis major	2
		right frontalis outer	
left eyebrow	raise	left frontalis major	
		left frontalis outer	
right eyebrow	raise frown	right frontalis major	2 + 4
		right frontalis outer	
		right secondary frontalis	
		right lateral corigator	
left eyebrow	raise frown	left frontalis major	
		left frontalis outer	
		left secondary frontalis	
		left lateral corigator	
right eyebrow	oblique	right frontalis inner	1
left eyebrow	oblique	right frontalis inner	
left lip corner	raise	left zygomatic major	12
right lip corner	raise	left zygomatic major	

(AUs) defined by Facial Action Coding System (Ekman and Friesen, 1978). A dictionary is used to define the relation between the facial component actions and Abstract Muscles. A facial component action describes an action in one part of the face and it can be seen as a AU or a group of AUs. Table 5.2 shows some examples of relations between facial component actions, Abstract Muscles and AUs.

5.4.6 Psyclone: Message-based Middleware

In Psyclone (Thórisson, 2004; Thórisson et al., 2004) a number of blackboards can be created and each module can register to these blackboards, either to post or to receive particular types of messages. Modules can post messages at any time. When a message arrives at a blackboard a wake up message is sent to each module which has registered to this blackboard and to this type of message. A wake up message contains the message that has been sent to the blackboard. A module receiving a wake up message will process it, will act on its content, and then it might send its own message.

A system based on a modular architecture, a message passing mechanism, and blackboard technology, implemented through Psyclone, offers many advantages:

- Information is available to the relevant modules
  As an example in the current implementation, the Dynamic Emotion Representation sends at regular time intervals, to a particular blackboard, a message reporting the current emotion state. This information becomes available to any module that needs it, such as the *communicative function dictionary module* and the *emotional dictionary module*. In the future, this message could also be used by a module which could change the colour of the skin or the dynamics of muscular actions.

- The insertion of new modules is facilitated
  The use of a message passing system facilitates the integration of new modules. For instance, a module controlling the expressiveness of the character could receive and modify facial signal messages before they arrive to the *Facial Component Action Dynamic Representation.*

- The run-time debugging is facilitated
  Psyclone also provides a web interface that is used to see all the blackboards and messages passing through the system at run-time. This facility and the possibility to send messages manually make the debugging easier and enables programmers to test each module separately.

- The reuse of the code is facilitated
  A module developed to be connected to Psyclone has an interface defining the services provided by this module. Such a module can easily be integrated in other applications by people who use Psyclone. The simplicity of use and the programming language independence of Psyclone make it a great tool for the creation of a library of connectable modules. People

could use this library to create more complex systems and participate in the effort of building intelligent virtual characters.

5.5 Timing and Synchronisation

Emotional expressions due to emotional impulses are not synchronised with the speech because they are reactions to emotional events, which appear from the internal and external environments of the virtual actor. Emotional impulse tags can be introduced in the text that should be spoken by the virtual actor but they describe unplanned events from the virtual actor point of view.

Communicative functions are supporting the speech and for this reason they should be synchronised with it. Visemes, which are the visual counterparts of phonemes, must also be synchronised with the speech because they represent visual mouth shapes that produce the sound of the speech.

5.5.1 Synchronisation of Communication Functions

Communicative functions are tightly synchronised with the speech and this is a reason why the communicative function tags are found within the text that should be spoken. In fact it is not exactly the communicative functions that are synchronised with the speech but their facial signals. The relation between a communicative function and time is the duration for which it is expressed, where the relation between a facial signal and time is more complex. A facial signal is defined as a list of facial component actions, and the relations between movements and time, such as speed and acceleration of the movement, are important, they are the signature of the expressed meanings.

The relation between facial signals and time is described by four durations: a *delay*, an *attack*, a *sustain*, and a *decay* duration. Examples of a relation between facial signal and time can be seen in Figure 5-8. This Figure shows the duration of a facial meaning and its corresponding facial signal with two different synchronisations.

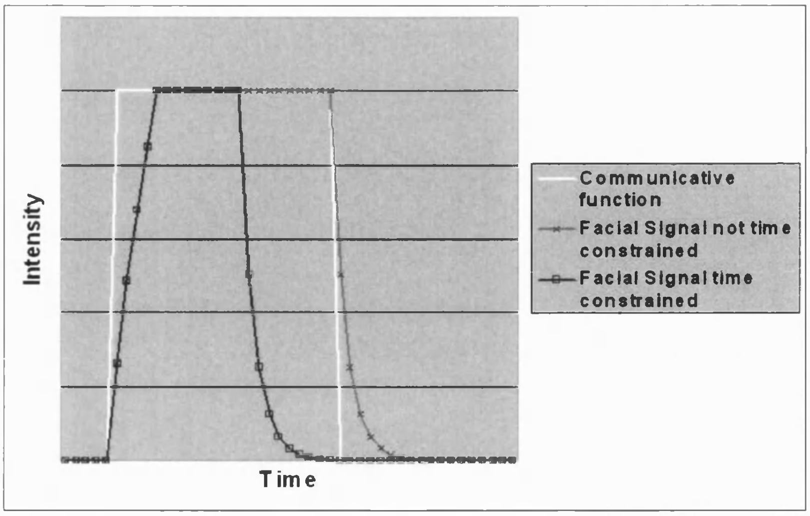

Figure 5-8: Two different synchronisations for a facial signal, in relation to the corresponding communicative function.

During the preparation of animation scripts, the durations of a communicative function are computed from the position of its tag within the text and the timing information of the phonemes corresponding to this text. The starting time of a communicative function is given by the starting time of the first phoneme met inside the corresponding tag. The duration of a communicative function is determined by the cumulative time length of the phonemes present between the beginning and the end of the corresponding communicative function tag.

When animation scripts are played communicative functions are transformed into facial signals bringing synchronisation issues: *should the attack and the decay durations of a facial signal be part of the duration of the communicative function, or should they be added to that duration?* Due to the real-time nature of this part of the system, facial signals are selected only at the time when they should be displayed. Therefore the attack duration and decay duration values of the signal are not known before the time at which the facial signal should be displayed. In consequence it is not possible to have the attack duration of the facial signal before the beginning of the communicative function. The end time of a facial signal can be changed at the "same" time as it is displayed so it is possible to include the decay duration in the duration of the communicative function or to concatenate it at the end. For facial signals tightly synchronised with words, such as the communicative function *emphasise*, it makes sense to include the decay duration within the boundaries of the words, e.g. in the boundaries of the communicative function. For other signals, the synchronisation could depend on the context in which they are used, such as the position in the sentence. For instance, the facial signal corresponding to the communicative function *question marker* might carry on over the boundary of the sentence. An example of two different synchronisations for a facial signal is shown on Figure 5-8. It should be emphasised that the facial signal curves on this figure describe the movement of only one facial part, i.e. one facial component action, and not the complete facial signal. EE-FAS allows a choice of which synchronisation should be used on a facial signal basis, by setting a parameter called `constrained_by_time`, in the XML dictionary described in section 5.4.4.

5.5.2 Synchronisation of Visemes

The synchronisation of visemes with sound is a topic of research by itself and not the topic of this thesis, therefore a simple mechanism has been implemented.

Visemes are the visual counterparts of phonemes and in the EE-FAS the duration of a viseme is the same as the duration of the corresponding phonemes. In this system the correspondence between visemes and phonemes is one-to-one.

For simplicity, the attack and decay durations of visemes are fixed, and the sustain duration varies as a function of the duration of the phoneme. The decay duration is added to the duration of the phoneme which means that two consecutive visemes will blend together. This method produce a more natural movement than to position the mouth at the neutral position between each viseme. The transition between two visemes is based on a mechanism limiting the speed of the Abstract Muscle movements, limiting the maximum distance between two contraction values within a certain amount of time.

Other models use coarticulation to synchronise visemes with speech such as the model described by Cohen and Massaro (1994).

5.6 Conclusion

This Chapter described the implementation of the Emotionally Expressive Facial Animation System (EE-FAS), which is based on facial meanings to control the expression of the face.

Existing facial animation systems using facial meanings to produce animations focus on either emotional expressions or on communicative acts. Generally, emotional meanings are included within the communicative meanings. In this thesis, it is argued that the facial expressions should be produced by two different processes: communicative and emotional. Facial expressions communicating something about emotions are visually different if they are due to emotional events than if due to a communication process. A new categorisation of facial meanings has been built for the EE-FAS from the descriptions of Ekman (1979) and Bavelas and Chovil (1997).

The EE-FAS displays Ekman's emotional expressions in response to emotional events to communicate clear emotional messages. These emotional expressions are not used during the speech, they can even stop the speech. In contrast, the EE-FAS uses communicative functions during the speech. To extend the facial vocabulary of the virtual actor, it displays different facial signals corresponding to one communicative function. The facial signals are selected according to the emotional state of the virtual actor.

A Dynamic Emotion Representation is introduced in the architecture of the EE-FAS to represent the emotional state of the virtual actor. The DER enables the EE-FAS to produce different animation according to the current state of the DER. It also represents the emotional history and the emotional momentum of the virtual actor and enables the EE-FAS to produce coherent facial expressions over time. Screen-shots of animation scripts played in different emotional contexts are presented and discussed to show these properties.

The implementation of the EE-FAS is based on a modular architecture, using a message passing mechanism for the communication between modules. The transformations between each Level of Control and their customisable definitions make the system flexible; able to animate different types of human-like virtual actors.

Part III

Perception of Facial Component Actions

Chapter 6

Perception of Facial Component Actions

6.1 Introduction

Previous chapters presented the design and the functionalities of the Emotionally Expressive Facial Animation System (EE-FAS) developed for this thesis. This chapter is concerned with the evaluation of the videos produced by the EE-FAS.

This Chapter describes an experiment that is used to show that the videos produced by the EE-FAS can communicate the same meanings as human facial expressions. Smith, Scott and Russell suggest from experimental results that movements of individual facial parts, called here facial component actions, communicate meanings and influence people's perception of the displayer's emotional state (Smith and Scott, 1997; Russell, 1997). For instance, an *eyebrow frown* movement is interpreted as a state of *displeasure* and a *lip corner raise* movement is interpreted as a state of *pleasure*. If facial movements are produced by the EE-FAS can be interpreted in the same fashion, it would show that the EE-FAS can create videos communicating interpretable facial meanings.

This experiment focuses on the interpretation of movements of particular facial parts, in contrast to full-face configurations such as the universally recognised expressions (Ekman, 1992; Izard, 1971). The EE-FAS does not display universally recognised emotional facial expressions during speech, it displays only movements of facial parts. This characteristic could mean that the EE-FAS might not be able to communicate emotions while speaking, which would cast a doubt on the EE-FAS's architecture. The experiment should make clear that emotions can be communicated through the displays of particular facial component actions.

Most of the experimental data are concerned with the meanings communicated by individual facial component actions. The experiment presented in this Chapter explores the interpretation of combinations of facial component actions. Particularly the combination of a smile and eyebrow frown, which is used in animation to represent nasty characters. To measure the effects of such combinations, the experiment uses the dimensions *Insincere/Sincere* and *Unfriendly/Friendly* as well as emotional dimensions.

The outlines of the experiment are presented in Section 6.2, and the experiment by itself, including independent and dependent variables, general and working hypothesis, is described in Section 6.3. The data is analysed in Section 6.4, and the results are presented in Section 6.5.

6.2 An Experiment to Evaluate Animations

The Emotionally Expressive Facial Animation System is based on a modular architecture and its design is inspired by psychology studies of emotions and facial expressions. As discussed in other Chapters, the EE-FAS is different from the existing systems mainly due to the use of a Dynamic Emotion Representation (DER) to select facial signals corresponding to communicative functions and to produce emotional expressions. With regard to other systems, the EE-FAS is based on a different categorisation of facial meanings.

Previous Chapters shown that the DER developed for this thesis maintains emotion momentums and that the EE-FAS can produce different animations according to the emotional state represented by the DER. This Chapter presents the evaluation of the animations produced by the EE-FAS through an experiment. This experiment aims to verify that the use of different facial movements is interpreted by people as communicating different meanings.

It is argued by many researchers that movements of facial parts communicate their own meanings (Smith and Scott, 1997; Russell, 1997). Table A.1, copied from (Russell, 1997, Table 13.1, p. 301), presents how facial actions influence people's perception of a person's state on the dimensions *pleasure* and *arousal*. Smith and Scott list facial component actions involved in emotional expressions and the meanings communicated by facial component actions (Smith and Scott, 1997). These descriptions are summarised in Tables A.2 and A.3 which are respectively copies of Tables 10.2 and 10.2 from Smith and Scott (1997).

Using the EE-FAS, it is possible to control individual facial component actions and to design an experiment testing whether people can interpret meanings from individual facial component actions on a synthetic face. Tables A.1, A.2 and A.3 are used as references to design such an experiment and to validate the results.

This experiment shows videos, produced by the EE-FAS, involving only a selected set of facial movements. All the videos are based on the same text, producing the same visemes which are visual counterparts of phonemes, and the same facial signals corresponding to the communicative functions. The exception is a communicative function expressing the emotional state of the virtual actor which is displayed at the same time as other facial signals. This communicative function is expressed by different facial signals according to the emotional state of the virtual actor; it is the effects of these facial signals that the experiment tries to measure. For instance, the virtual actor shown in the videos could display a smile or an eyebrow frown while talking, depending on its emotional state. The experiment would measure the influences of these facial component actions on people's perception of the virtual actor's level of happiness. The experiment measure the effects of individual facial component actions in contrast to full-face configurations.

The effects of these facial signals can be measured on many different dimensions. In the first place, it is interesting to measure their influences on *emotional* dimensions due to the use of the emotional state of the virtual actor to produce the videos. A correlation should be found between the emotional state controlling the facial expressions and the perceived emotional state. This should also prove that the EE-FAS can communicate emotional messages during speech. To measure the emotional state of the virtual actor, two emotion models are used. The first model is a categorical discrete emotion model composed of basic emotions such as Happy, Angry, Sad and Disgusted (Ortony and Turner, 1990; Edwards, 1999; Ekman, 1999a). Using these categories, the subjects are asked to rate how this type of emotion is felt by the virtual actor. For instance, how happy is the virtual actor on a scale 1 to 5. The categorical model is interesting because the basic emotions are related to typical facial expressions, therefore the subjects should be able to rate the virtual actor from its facial expression using these categories. The second model is inspired by the model described by Russell (1997) and shown in Figure 6-1. The dimensions Unpleased/Pleased and Lethargic/Energetic are used to describe an emotional space similar to the one shown in Figure 6-1. Using these two dimensions it is possible to rate the emotional state of the actor. The advantage of this model is the mapping of emotion categories, such as Happy and Sad, onto the emotional space defined by the two dimensions. From the experimental data gather by this experiment, a correlation should be found between the two models. For instance, if the virtual actor is rated highly on the category happy, its score on the dimensions Unpleased/Pleased and Lethargic/Energetic should correspond to the position of this category.

In addition to the emotional state of the virtual actor, it is interesting to see if facial component actions influence people's perception of the character on dimensions such as *friendliness* and *sincerity.* These dimensions have been chosen to verify the hypothesis that combinations of facial component actions having opposite meanings, such as eyebrow frown and smile reflecting respectively anger and happiness, would affect negatively people's perception of the character's sincerity. This hypothesis comes from that this type of expression, e.g. smile and eyebrow frown, is due to the masking of an expression of anger with a smile. The act of masking felt emotions might be interpreted by people as insincere. This type of expression is also used in animated characters playing the role of a nasty character. Examples of this type of expressions are shown in Figures 6-2 and 6-3. Figure 6-2 shows snapshots of Gollum from the film the Lords Of The Rings: The Two Towers (2002), when he talks to himself, taking in turns the interpretation of a gentle character and a nasty violent character. Picture 6-2(a) is taken when Gollum is driven by the gentle personality, where picture 6-2(b) shows Gollum driven by a nasty personality. In these two pictures Gollum smiles, but when the nasty personality is expressed the smile is accompanied with

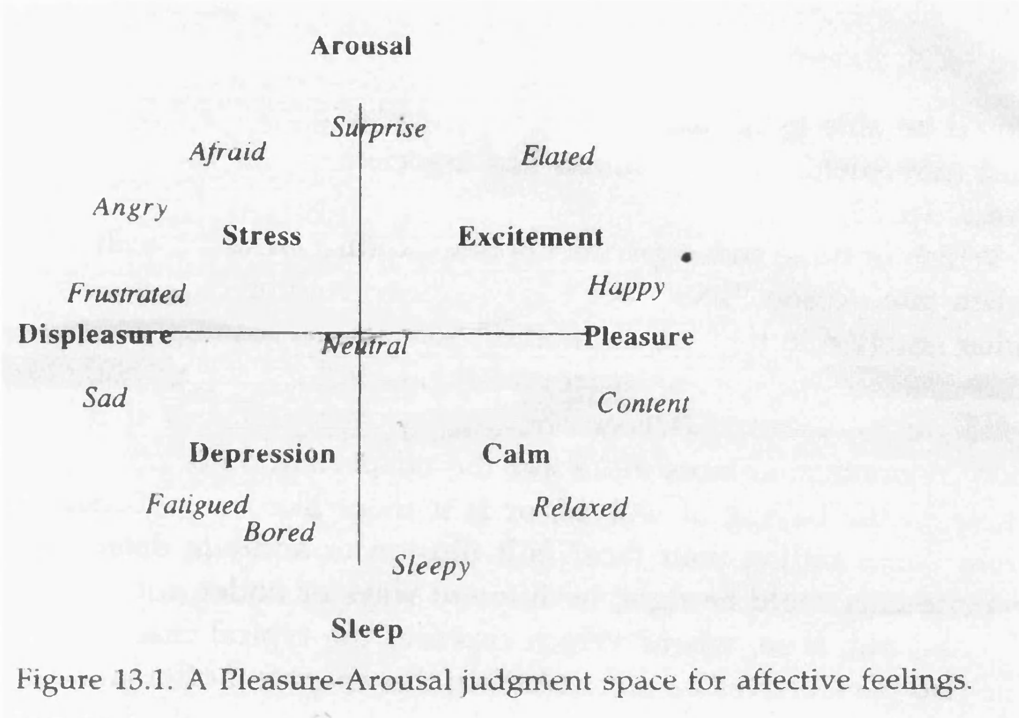

Figure 6-1: Emotional space defined by two dimensions: Displeasure/Pleasure (Russell, 1997, Fig 13.1, p. 299)

an eyebrow frown. Another example is the Figure 6-3 showing a doll called Max which is the main character of a children book by Sendak (1963). In this book, Max is a naughty child and to represent this characteristic the doll's face shows an expression combining a smile with an eyebrow frown.

Finally, it would be interesting to see if people can be influenced by subtle differences between facial expressions. For instance, the differences between a genuine and a fake smile is defined as actions of muscles around the eyes during a genuine smile. The current implementation of the Abstract Muscle model does not support these types of actions but, as shown in Table A.2, happiness is communicated through a facial signal composed of *raised lower eyelids*, *lip corners raised* and an *open mouth*. Therefore two types of smile are shown in this experiment, a genuine smile involving all the facial component actions previously mentioned and a fake smile composed of only the facial component action *lip corners raise*. This experiment verifies whether people can perceive subtle differences and whether

(a) Gollum driven by a gentle personality.

(b) Gollum driven by a nasty violent personality.

Figure 6-2: The character Gollum in the film The Lords Of The Rings: The Two Towers (2002)

Figure 6-3: Max, a naughty child from Sendak (1963).

these differences influence their perception of the virtual actor.

6.3 Design of the Experiment

This section presents the design of the experiment by defining *independent* and *dependent variables*, stating the *hypotheses* that the experiment tries to prove, and describing the setup of the experiment. Finally, the creation process of the videos used for this experiment is explained at the end of this section.

6.3.1 Experimental Variables

For an experiment two types of variables need to be defined, the *independent* and *dependent variables.* The independent variables are the features changing across conditions and the dependent variables are used to measure the effect of the independent variables.

Independent Variables and Conditions

The independent variables in this experiment are facial components actions. Four independent variables are defined:

- Eyebrow frown (AU4)
- Eyebrow oblique (AU1)
- Lip corners raise (AU12)
- Raise lower eyelid + open mouth (AU6, AU7, AU26 and AU27)

Using these four independent variables, nine combinations are used to create the videos for this experiment. Typical facial expressions corresponding to these combination are show Figure 6-4.

1. No facial component actions (neutral condition, control condition)
2. Eyebrow frown (AU4)
3. Eyebrow oblique (AU1)
4. Lip corners raise (AU12)
5. Eyebrow frown (AU4) + Lip corners raise (AU12)
6. Eyebrow frown (AU4) + Lip corners raise (AU12) + Raise lower eyelid + open mouth (AU6, AU7, AU26 and AU27)

7. Eyebrow oblique (AU1) + Lip corners raise (AU12)

8. Eyebrow oblique (AU1) + Lip corners raise (AU12) + Raise lower eyelid + open mouth (AU6, AU7, AU26 and AU27)

9. Lip corners raise (AU12) + Raise lower eyelid + open mouth (AU6, AU7, AU26 and AU27)

Each one of these combinations, except the control condition, is displayed with two intensities: 40% and 80% of the maximal intensity of the facial component actions. These two percentages have been chosen because they cover the range of intensities without being extremes. In total *17 conditions*, e.g. 17 videos, are produced for this experiment.

Dependent Variables

To measure the influences of the independent variables, seven dependent variables are defined.

A. Unfriendly/Friendly

B. Insincere/Sincere

C. Lethargic/Energetic

D. Unpleased/Pleased

E. Happy

F. Sad

G. Angry

The variables Unfriendly/Friendly and Insincere/Sincere have been chosen to measure the effect of the independent variables on dimensions other than emotional. The variables Lethargic/Energetic and Unpleased/Pleased are used to define an emotional space relatively similar to the model described by Russell

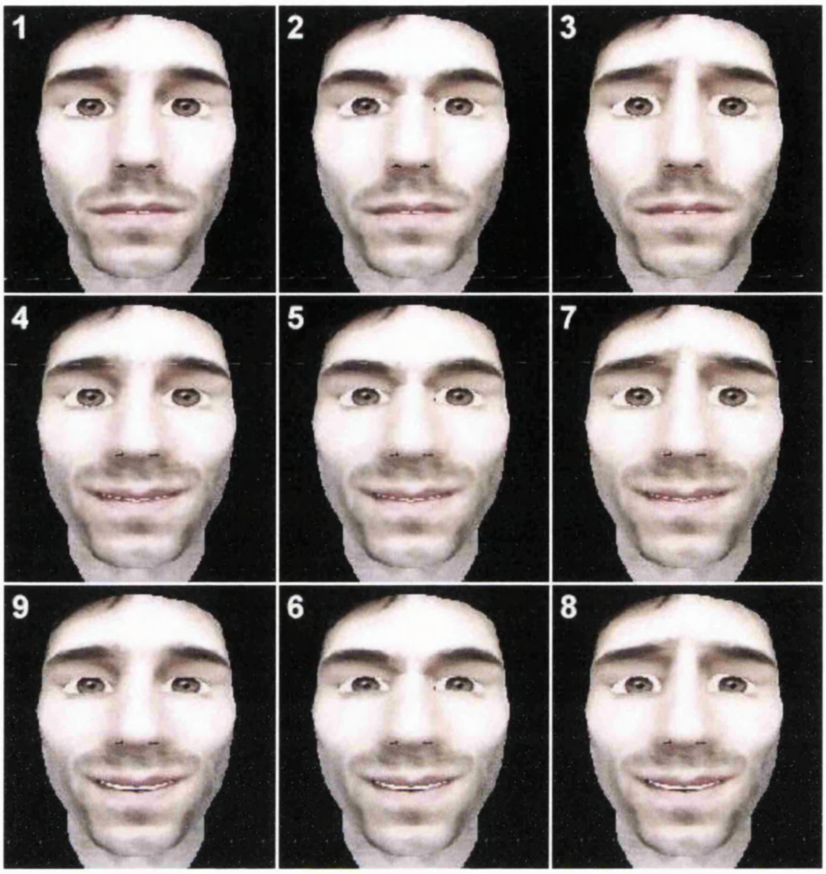

Figure 6-4: Typical facial expressions representing the combinations of facial component actions used for the experiment. The numbers correspond to the list of combinations in Section 6.3.1.

(1997). The dimensions used by Russell are not well adapted to ask questions about the state of a third person, such as the virtual actor. It is assumed that the dimensions Lethargic/Energetic and Unpleased/Pleased are close enough to the dimensions Displeasure/Pleasure and Sleep/Arousal to represent the same emotional space. The dimensions Happy, Sad and Angry have been selected because the facial component actions displayed during this experiment are involved in the facial expression of these emotions (Smith and Scott, 1997). These emotions can also be mapped onto the emotional space described by Russell.

The dimensions Unfriendly/Friendly, Insincere/Sincere, Lethargic/Energetic and Unpleased/Pleased are measured on a scale of -3 to 3. For these dimensions a

7 points scale is used because they are composed of two extremes. This type of scale is commonly used in questionnaire due its statistical reliability and it also permits to distinguish between differences within both extremes.

The dimensions Happy, Sad and Angry are measured on a scale of 0 to 4. In this case a 5 points scale is chosen because the dimensions are composed of only one term and it is difficult for the subjects to distinguish between small differences.

6.3.2 Experimental Hypotheses

The formulation of hypotheses is used to define the expected conclusions from the experimental data.

General Hypotheses

Four general hypotheses are defined:

General Hypothesis 1 (GH_1): Facial component actions, such as those related to eyebrow frown (AU4) and lip corner raise (AU12), communicate meanings and influence people's perception of a virtual character on dimensions such as Unfriendly/Friendly, Unpleased/Pleased, Happy and Angry.

General Hypothesis 2 (GH_2): Facial expressions involving facial component actions that have opposite meanings, such as eyebrow frown (AU4) and lip corners raise (AU12), influence negatively people's perception of the character's sincerity.

General Hypothesis 3 (GH_3): The meaning of a facial expression is a combination of the meanings of its part.

General Hypothesis 4 (GH_4): Subtle differences in facial expressions, such as lower eyelid raised, can influence people's perception of the virtual character.

Working Hypotheses

To draw conclusions from the experimental data, working hypotheses, which are more closely related to the variables defined for the experiment than the general hypotheses, are defined. The following working hypotheses describe the expected results:

- *Working Hypothesis 1 (WH_1)*: Eyebrow frown movements are interpreted as a sign of unfriendliness and they influence negatively people's perception of the virtual actor on the dimension Unfriendly/Friendly.
- *Working Hypothesis 2 (WH_2)*: Lip corner raise movements are interpreted as a sign of friendliness and they influence positively people's perception of the virtual actor on the dimension Unfriendly/Friendly.
- *Working Hypothesis 3 (WH_3)*: Facial expressions involving lip corner raise and eyebrow frown movements at the same time are interpreted as a sign of insincerity and they influence negatively people's perception of the virtual actor on the dimension Insincere/Sincere.
- *Working Hypothesis 4 (WH_4)*: Facial expressions involving lip corner raise, lower eyelid raise and open mouth movements at same time, influence more positively the dimension Insincere/Sincere than facial expressions involving only lip corner raise movements.
- *Working Hypothesis 5 (WH_5)*: Facial expressions involving lip corner raise, lower eyelid raise and open mouth movements at same time, influence more positively the dimension Unfriendly/Friendly than facial expressions involving only lip corner raise movements.
- *Working Hypothesis 6 (WH_6)*: Eyebrow frown movements are interpreted as a sign of an unpleased state and they influence negatively on the dimension Unpleased/Pleased.
- *Working Hypothesis 7 (WH_7)*: Eyebrow frown movements are interpreted as a sign of an energetic state and they influence positively on the dimension Lethargic/Energetic.
- *Working Hypothesis 8 (WH_8)*: Eyebrow oblique movements are interpreted as a sign of unpleased state and they influence negatively on the dimension Unpleased/Pleased.

- *Working Hypothesis 9 (WH_9)*: Eyebrow oblique movements are interpreted as a sign of a lethargic state and they influence negatively on the dimension Lethargic/Energetic.
- *Working Hypothesis 10 (WH_{10})*: lip corners raise movements are interpreted as a sign of a pleased state and they influence positively on the dimension Unpleased/Pleased.
- *Working Hypothesis 11 (WH_{11})*: lip corner raise movements are interpreted as a sign of an energetic state and they influence positively on the dimension Lethargic/Energetic.
- *Working Hypothesis 12 (WH_{12})*: lip corner raise movements are interpreted as a sign of a happy state and they influence positively on the dimension Happy.
- *Working Hypothesis 13 (WH_{13})*: Facial expressions involving lip corner raise, lower eyelid raise and open mouth movements at same time, influence more positively the dimension Happy than facial expressions involving only lip corner raise movements.
- *Working Hypothesis 14 (WH_{14})*: eyebrow oblique movements are interpreted as a sign of a sad state and they influence positively on the dimension Sad.
- *Working Hypothesis 15 (WH_{15})*: eyebrow frown movements are interpreted as a sign of an angry state and they influence positively on the dimension Anger.

6.3.3 Experiment Setup

For this experiment, all the 17 videos are shown to each subject and after each video the subject is asked to rank the virtual actor shown in the video on the seven dimensions described in Section 6.3.1. The videos are displayed in a randomised order.

The experiment was presented as web pages connected to a database to display the videos and the questions and to gather the subject's answers. The first pages introduce the experiment, then the following pages gather data about the subjects, such as age, gender, and nationality. Before the subject starts the experiment the series of questions asked after each video is presented. Each question is presented with several radio buttons, one for each possible answer. The subjects can give only one answer for each question. Examples of a question are shown in Tables 6.1 and 6.2, but the numerical values of the answer are not shown to the subjects.

Table 6.1: Example of a 7 point scale question.

Mr. Smith was:

Very Friendly	Friendly	Little Friendly	Neutral	Little Unfriendly	Unfriendly	Very Unfriendly
3	2	1	0	-1	-2	-3

Table 6.2: Example of a 5 point scale question.

Mr. Smith was:

Not Happy	Slightly Happy	Happy	Very Happy	Extremely Happy
0	1	2	3	4

6.3.4 Creation of the Videos for the Experiment

Each video is created from the following text:

```
<text>
 <comm_func name="personal reaction" max_nb_fd="2" intensity="40" />
  <comm_func name="PauseMarker" intensity="100" />
  Good morning and
  <comm_func name="PauseMarker" intensity="100" />
  welcome to the Department of
  <comm_func name="PauseMarker" intensity="100" />
  Computer Science,
  <comm_func name="PauseMarker" intensity="100" />
  at the University of Bath.
  <comm_func name="PauseMarker" intensity="100" />
```

```
  </comm_func>
</text>
```

The tags `<comm_func name="PauseMarker" intensity="100" />` are used to control the blinking of the virtual actor and they are the same in each video. The only differences are due to the intensity of the communicative function *personal reaction* and the emotional state represented by the DER. For a particular emotional state, a signal is associated with the communicative function *personal reaction*. For this experiment, the facial signals are reduced to only one or few facial component actions. For instance, in the emotional state of happiness, the lip corners raised movement is shown; in the emotional state of anger the eyebrow frown movement is shown; in the emotional state of sadness the eyebrow oblique movement is shown. Because the attribute *max_nb_fd* has the value 2, if two emotions are present, two facial signals are displayed. For instance, in the emotional state of anger and happiness, the facial signals eyebrow frown and lip corners raised are displayed. This is not a good example of how to use the EE-FAS because it is a very improbable emotional state but for the needs of the experiment this emotional state can be enforced. So by composing emotional states, the EE-FAS can produce animations displaying the combinations of facial component actions described in Section 6.3.1. To create the facial combination numbered 6, 8 and 9, the definition of the facial signal displayed in the context of happiness has been modified. In these combinations, the smile is composed of the facial component actions *lip corners raise*, *raise lower eyelid* and *open mouth*. Finally, for each emotional state, the intensity of the communicative function *personal reaction* takes the values 40% and 80%.

6.4 Analysis of the Experimental Data

The analysis of data gathered through the experiment is carried out in two steps. In the first place, the influences of the principal facial component actions, such as an eyebrow frown and lip corners raise, are analysed. These are the combinations numbered 2, 3, 4, 5 and 7 described in Section 6.3.1. Afterwards, the differences due to more subtle facial component actions, such as lower eyelid raise, are discussed. These combinations are numbered 6, 8 and 9 in Section 6.3.1. In total, 60 subjects, 31 males and 29 females, participated in this experiment. All the experimental data are presented in a graph format plotting the mean of all the answers for each conditions.

6.4.1 Influences of the Main Facial Component Actions

In this section, all the comparisons between conditions are based on the results gathered with the facial configurations at the higher intensity (80%). This is due to the fact that most of the curves are monotone and therefore differences are more significant in the extreme cases. However, exceptions exist, the facial component action *eyebrow frown* at a low intensity (40%) has less influence on people's perception than other facial component actions at the same intensity. Where the relations between the influences of certain facial component actions and their intensities are almost linear, such as in Figures 6-5 and 6-7, the relation between the influence of an *eyebrow frown* and its intensity is not linear, such as in Figure 6-6. Therefore, in some cases when an *eyebrow frown* is combined with another facial component action at low intensity (40%), the subjects are influenced as if the *eyebrow frown* was not displayed, such as in Figure 6-5 and 6-9. The analysis of the data gathered by the facial signals at an intensity of 80% should be less affected by this problem because all the facial component actions should be perceived.

This section only analyses the influences of the combinations of the following facial component actions: *eyebrow frown*, *eyebrow oblique* and *lip corners raise*. To simplify the description of the analyses, *eyebrow frown*, *eyebrow oblique* and *lip corners raise* are called respectively from now on *frown*, *oblique* and *smile*. This section is concerned with the combination numbered 2, 3, 4, 5 and 7 in Section 6.3.1.

The condition showing non facial component action related to the emotional state, e.g. the control condition numbered 1, and the conditions showing facial component actions at an intensity of 80% are compared. These comparisons are used to draw conclusions on the influences of a particular facial component action or a particular combination of facial component actions. The results for facial component actions at an intensity of 80% are also compared between each other to evaluate the differences in influence between facial component actions.

To verify if differences between conditions are significant, the paired-sample t-test is used and the results are highly significant when $p \leq 0.001$.

Analysis on the Emotional Dimensions

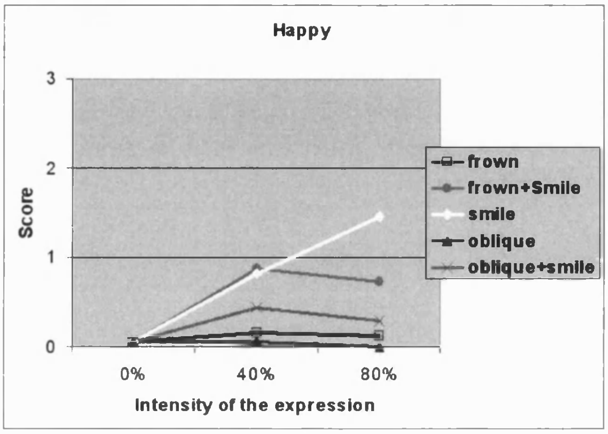

Figure 6-5: Averages of the answers for the dimension happy.

Figure 6-5 shows that a smile is interpreted by people as a sign of happiness. The combination of a smile and frown is also interpreted as a sign of happiness but with less intensity than just a smile. The differences between neutral and smile conditions, and neutral and smile+frown conditions are highly significant. The difference between smile and smile+frown conditions is also highly significant, therefore a frown reduces the level of happiness communicated by a smile. The difference between smile+frown and smile+oblique is also highly significant, showing that an oblique movement reduces the influence of a smile even more than an eyebrow frown movement.

Figure 6-6 shows that a frown movement is interpreted by people as a sign of anger, even if this frown is combined with a smile. The differences between the neutral and frown conditions, and neutral and frown+smile conditions are highly significant.

Figure 6-7 shows that oblique eyebrows are interpreted by people as a sign of

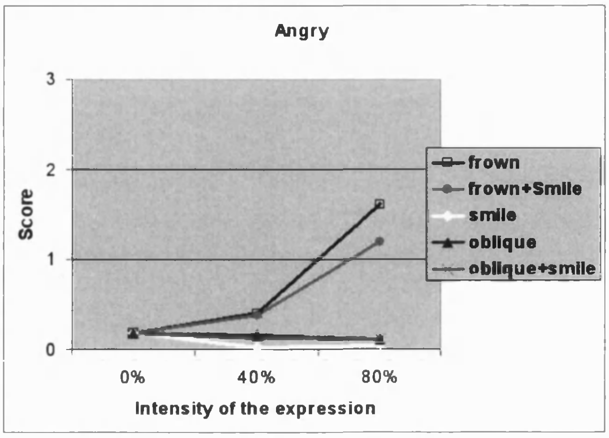

Figure 6-6: Averages of the answers for the dimension angry.

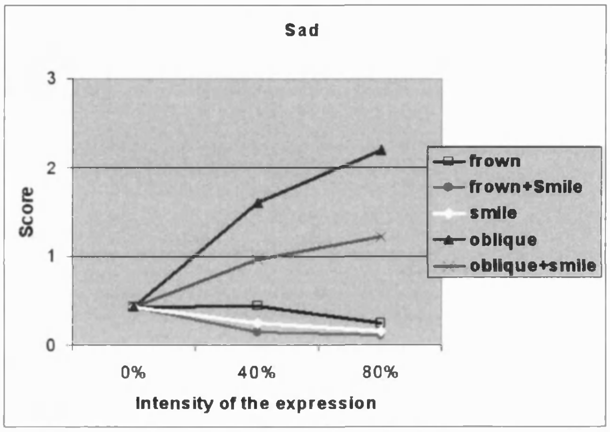

Figure 6-7: Averages of the answers for the dimension sad.

sadness, even combined with a smile. The differences between the neutral and oblique conditions, and neutral and oblique+smile conditions are highly significant. The combination of oblique eyebrows and smile reduce (highly) significantly the communication of sadness.

Figure 6-8 shows people's perception of facial component actions on the dimension Lethargic/Energetic and Unpleased/Pleased. The emotional space described by these two dimensions is closely related to the emotional space shown in Figure 6-1. It should be emphasised that the control condition, called neutral in Figure 6-8, is not at the origin of the emotional space. By moving the origin of the space

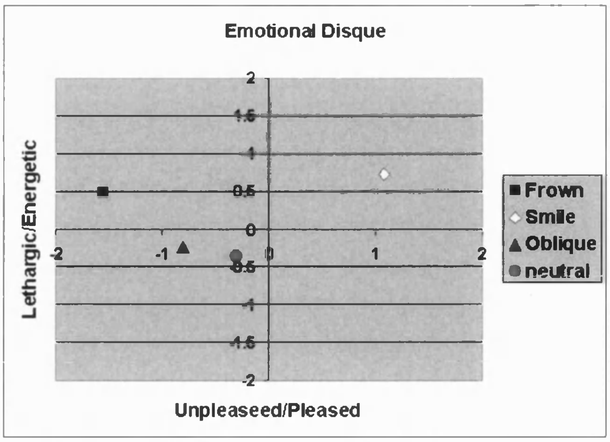

Figure 6-8: Averages of the answers for the dimensions lethargic/energetic and unpleased/pleased

to the point representing the neutral condition and mapping the two spaces, Figures 6-1 and 6-8, it is possible to see how a particular facial component action influences perception of the virtual actor's emotional state.

For the dimensions *Unpleased/Pleased* and *Lethargic/Energetic* the scale is -3 to 3. From the experimental data the Neutral condition has the coordinates (-0.3,-0.3667). Table 6.3 shows: the coordinates of the facial component actions in the original emotional space; the coordinates in the translated emotional space, e.g. moving the origin to the neutral condition point; the coordinates on a scale -4 to 4 to compare them to the results presented in Table A.1; and the coordinates shown in Table A.1. Table A.1 does not show the result for *lip corner raise* so this facial component action is compared with the results of *cheek raised* which is the closest to the *lip corner raise* movement.

Coordinates from experimental data which are in the right eighth of the emotional disc are considered to be matching the correct interpretation. Therefore, from the results shown in Figure 6-8 and Table 6.3, a *frown* movement can be mapped to an emotional state of *anger* and a *smile* can be mapped to an emotional state of *happiness*. The *eyebrow oblique* movement is not recognised as a reflection of the expected emotional state of sadness due to its positive score on the dimension Lethargic/Energetic.

Table 6.3: Comparison of the results obtained on the dimensions *Unpleased/Pleased* and *Lethargic/Energetic* with the results shown in Table A.1.

Facial Component Action	Original emotional space	Translated emotional space	-4 to 4 scale	Result from Table A.1 (-4 to 4 scale)
Frown	(-1.55,0.483)	(-1.25,0.85)	(-1.66,1.133)	AU4 (-2.00,1.23)
Oblique	(-0.8,-0.233)	(-0.5,0.1337)	(-0.666,0.178)	AU1 (-1.92,-0.85)
Smile	(1.083,0.733)	(1.383,1.0997)	(1.844,1.466)	AU6 (Cheek raised) (1.46,0.73)

Analysis of the Dimensions Unfriendly/Friendly and Insincere/Sincere

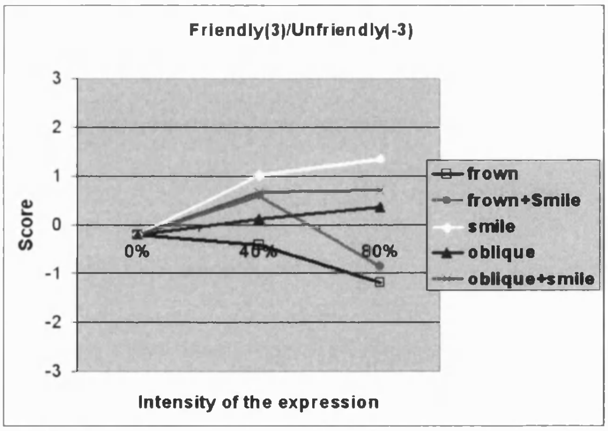

Figure 6-9: Averages of the answers for the dimension unfriendly/friendly.

Figure 6-9 shows that a smile is interpreted by people as a sign of friendliness and that a frown is interpreted as a sign of unfriendliness. The differences between the neutral and smile conditions, and neutral and frown conditions are highly significant. The facial component action *oblique eyebrows* and *oblique+smile* are interpreted as a sign of friendliness but with less intensity than just a smile. The differences between neutral and oblique conditions, and neutral and oblique+smile conditions are highly significant; and the differences between smile and oblique conditions, and smile and oblique+smile conditions are highly significant too. The combination of a frown and smile is interpreted as an unfriendly sign and this combination is diametrically opposite to the interpretation of a smile by itself.

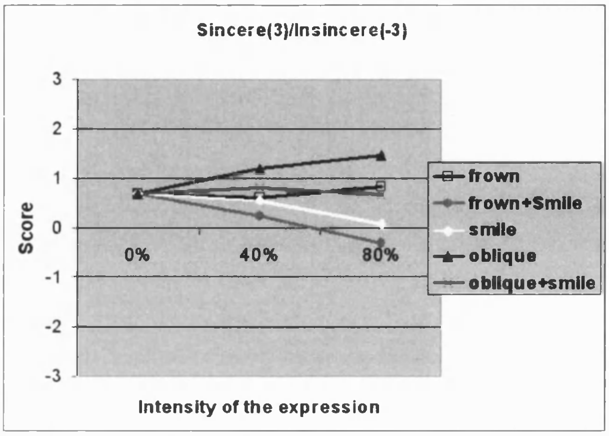

Figure 6-10: Averages of the answers for the dimension insincere/sincere.

Figure 6-10 shows that oblique eyebrows are interpreted as a sign of sincerity and that a high intensity smile reduces the perceived level of sincerity. The differences between neutral and smile conditions, and neutral and oblique conditions are highly significant. The smile+frown condition is interpreted as an insincere sign but the difference between smile and smile+frown condition is not statistically significant ($p=0.182$). Here, the low score of the smile condition is due to the fact that not all the components of a smile are displayed therefore it is not interpreted as a genuine smile. It is interesting to note the difference between male and female perceptions; for men the difference between smile and smile+frown conditions has a p value of 0.075, where the p value for females is 0.867. This means that females score the combination of smile+frown in almost the same way as a smile on the dimension insincere/sincere.

6.4.2 Influences of Subtle Differences in Facial Signals

This section analyses the conditions numbered from 5 to 9 to see whether subtle differences in facial signals, e.g. facial component actions *lower eyelid raise* and *mouth open*, could influence people's perception of the virtual actor. These two facial component actions are displayed only when the facial component action *lip corners raise* is displayed to try to simulate the differences between genuine and fake smiles. The comparisons are carried out only between facial signals differing due to the display or the non-display of *lower eyelid raise* and *mouth open*.

To verify if differences between conditions are significant, the paired-sample t-test is used. The results are borderline significant for $p \leq 0.05$, significant for $p \leq 0.01$ and highly significant for $p \leq 0.001$.

To simplify the description of the analyses, *eyebrow frown*, *eyebrow oblique*, *lip corners raise* and *lip corners raise, lower eyelid raise, open mouth* are called respectively from now on *frown*, *oblique*, *smile* and *complete smile*.

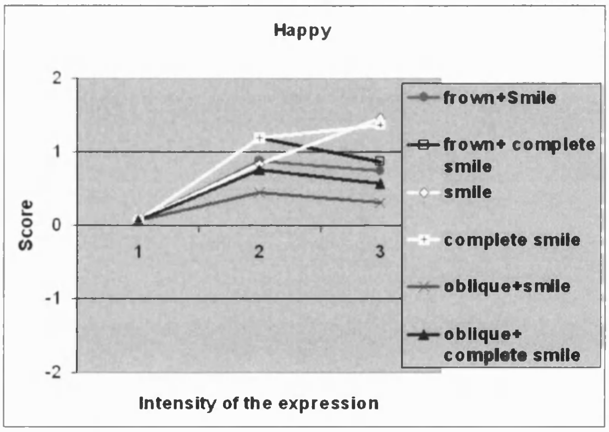

Figure 6-11: Averages of the answers

In Figure 6-11, the difference between a *smile* and a *complete smile*, both at an intensity of 40%, is borderline significant (p=0.031). In the same Figure, the difference between a *frown+smile* and a *frown+complete smile*, both at an intensity of 40%, is significant (p=0.01).

In Figure 6-12 the difference between a *frown+smile* and a *frown+complete smile*,

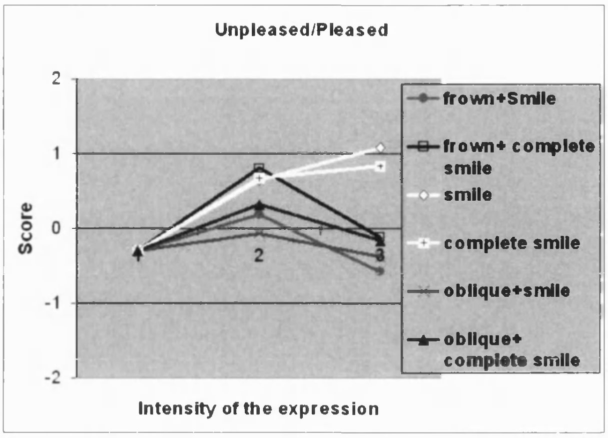

Figure 6-12: Averages of the answers

both at an intensity of 40%, is significant (p=0.006).

These results let us conclude that these subtle differences have influences only when the facial signals are displayed with a low intensity. The facial component actions *lower eyelid raise* and *mouth open* reduce the influences of displaying low intensity *eyebrow frown* on the dimensions *Happy* and *Unpleased/Pleased.*

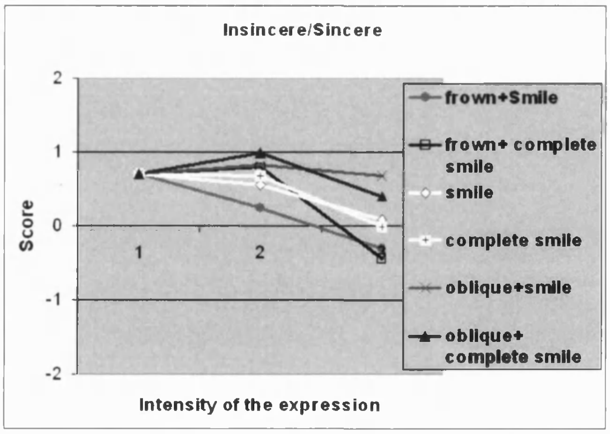

Figure 6-13: Averages of the answers

Figure 6-13 shows the difference between a *frown+smile* and a *frown+complete smile*, both at an intensity of 40%, is borderline statically significant (p=0.019) on the dimension *Insincere/Sincere.*

6.5 Presentation of the Results

6.5.1 Validation of the Hypotheses

Table 6.4 shows the statical analysis used to validate the working hypothesis defined in Section 6.3.2. Two conditions are needed to validate a hypothesis: The difference between the two conditions involved in the hypothesis should be significant; and the influence of the independent variable should be in the direction stated in the hypothesis. For instance, the hypothesis WH_{15} is not validated with an intensity of 40% because the difference between the two conditions is not significant, and it is not validated with an intensity 80% due to the negative value of the mean showing an influence opposite to the influence described by the hypothesis. If a hypothesis does not specify both conditions that should be compared, it means that the condition described in the hypothesis should be compared with the neutral condition. The sign $\times$ is used when a hypothesis is *not confirmed* and the sign $\surd$ is used when the hypothesis is *confirmed.*

Table 6.4: Validation of the Working Hypothesis

	Mean		t		Significance		Validation	
Int.	40%	80%	40%	80%	40%	80%	40%	80%
Hyp.								
WH_1	-0.216	-1.00	-1.897	-6.072	$p = 0.063$	$p \leq 0.001$	$\times$	$\surd$
WH_2	0.966	1.55	6.878	8.605	$p \leq 0.001$	$p \leq 0.001$	$\surd$	$\surd$
WH_3	-0.683	-1.017	-3.400	-4.088	$p \leq 0.001$	$p \leq 0.001$	$\surd$	$\surd$
WH_4	0.100	-0.233	0.511	-0.841	$p = 0.611$	$p = 0.404$	$\times$	$\times$
WH_5	0.117	-0.783	0.817	-3.492	$p = 0.417$	$p \leq 0.001$	$\times$	$\times$
WH_6	-0.267	-1.250	-2.583	-8.811	$p = 0.012$	$p \leq 0.001$	$\times$	$\surd$
WH_7	-0.367	0.850	-2.251	5.815	$p = 0.028$	$p \leq 0.001$	$\times$	$\surd$
WH_8	-0.317	-0.500	-3.385	-4.250	$p \leq 0.001$	$p \leq 0.001$	$\surd$	$\surd$
WH_9	-0.400	0.133	-2.560	0.806	$p = 0.013$	$p = 0.424$	$\times$	$\times$
WH_{10}	0.800	1.383	6.899	8.064	$p \leq 0.001$	$p \leq 0.001$	$\surd$	$\surd$
WH_{11}	0.417	1.100	2.634	6.092	$p = 0.011$	$p \leq 0.001$	$\times$	$\surd$
WH_{12}	0.550	1.400	6.846	11.091	$p \leq 0.001$	$p \leq 0.001$	$\surd$	$\surd$
WH_{13}	0.267	-0.533	2.206	-2.934	$p = 0.031$	$p \leq 0.005$	$\times$	$\times$
WH_{14}	1.167	1.767.	8.937	11.035	$p \leq 0.001$	$p \leq 0.001$	$\surd$	$\surd$
WH_{15}	0.217	1.433	3.023	12.475	$p \leq 0.005$	$p \leq 0.001$	$\surd$	$\surd$

The general hypothesis GH_1 has been proven to be true with the majority of facial component actions used in this experiment. Facial component actions influence people's perception of a virtual character. Subtle facial component actions, such as lower eyelid raise, also influence people but not in every condition.

The general hypothesis GH_2 has been proven to be true for the combination of *lip corner raise* and *eyebrow frown*, and not true for the combination *lip corner raise* and *eyebrow oblique*. Certain combinations of facial component actions expressing opposite emotional state are interpreted as *insincere*.

The general hypothesis GH_3 has been proven to be true. When facial movements are combined, most of the time, the meaning of a combination is a combination of the meanings of the parts; therefore mixing facial movements in different part of the face creates no new meaning. In the figures shown in this chapter, the meanings of the combinations are somewhere between the meanings of the individual facial movements. The exception is the combination of a smile and a frown which is interpreted as insincere. In figure 6-10, the meaning of this combination is not between the meanings of the individual facial movements. The facial component actions smile and eyebrow frown are not interpreted individually as a sign of insincerity; where as the combination of these two facial component actions is interpreted as a sign of insincerity. The difference between the condition involving only a smile, which has a lower score on the dimension *Insincere/Sincere* than an eyebrow frown, and the condition involving both a smile and an eyebrow frown is not statistically significant with this sample size.

The general hypothesis GH_4 has been proven to be true with facial signals of low intensity. With facial signals of low intensity, subtle differences in facial signal influence people's perception of a virtual actor.

The unexpected influences of the facial component action oblique eyebrow should be emphasised. Oblique eyebrow movements influence positively the dimension Friendliness/Unfriendliness and Sincere/Insincere.

6.5.2 Comparison with Existing Results

The data gathered by this experiment can be compared with existing results presented in Tables A.1, A.2 and A.3. This comparison is carried out on the emotional dimensions and does not concerned the combinations of facial component actions.

This experiment shows that individual facial component actions communicate their own meanings. According to the facial component action displayed, people's perception of the virtual actor changes.

In this experiment, the facial component action *eyebrow frown* (AU4) is interpreted as a sign of anger and a sign of an unpleased state. The interpretation of *eyebrow frown* as a sign of anger was expected due to the presence of this facial component action in the expression of anger, as described in Table A.2. The experimental results obtained for the dimension *Unpleased/Pleased* are similar to the results shown in Tables A.1 and A.3, where *eyebrow frown* influences negatively the dimensions *Pleasantness* and *Pleasure*, respectively. The influences of *eyebrow frown* on the dimensions *Lethargic/Energetic* and *Unpleased/Pleased* are similar to the influences on the dimensions *Arousal* and *Pleasure* reported in Table A.1. Therefore this facial component action is interpreted as a sign of anger in the emotional space defined by the dimensions *Lethargic/Energetic* and *Unpleased/Pleased*, as well as on the dimension *Angry*.

Through the data gathered by the experiment it is possible to conclude that the facial component action *lip corner raise* (AU12) is interpreted as a sign of happiness and a pleased state. The interpretation of *lip corner raise* as a sign of happiness was expected due to the presence of this facial component action in the expression of happiness, as described in Table A.2. The experimental results obtained on the dimension *Unpleased/Pleased* are similar to the results shown in Table A.3, where *lip corner raise* influences positively the dimension *Pleasantness.* The table A.1 does not provide any result regarding *lip corner raise.* The facial component action *lip corner raise* is interpreted as a sign of happiness on the dimensions *Unpleased/Pleased* as well as on the dimension *Happy.*

The data gathered through this experiment shows that *eyebrow oblique* movement (AU1) is interpreted as a sign of sadness on the dimension *Sad.* This result was expected due to the presence of this facial component action in the expression of sadness, as described in Table A.2. *Eyebrow oblique* scores negatively on the dimension *Unpleased/Pleased* reproducing the result presented in Tables A.1 and A.3. The influence of *eyebrow oblique* cannot be mapped to the emotional state of sadness in the emotional space described by the two dimensions *Lethargic/Energetic* and *Unpleased/Pleased.* The facial component action *eyebrow oblique* does not influence negatively people's perception on the dimension *Lethargic/Energetic*, which is a different result than the one presented in Table A.1. In the experiment, eyebrow oblique movement score low values on the dimensions Pleased/Unpleased and Lethargic/Energetic, in comparison to the result shown in Table A.1.

6.5.3 Correlation between Results and EE-FAS's Controls

The videos used in this experiment have been created by changing the emotional state of the virtual actor. For each emotional state, a particular facial component action is shown. The Emotionally Expressive Facial Animation System associated the emotional state of *sadness*, *happiness* and *anger* to the facial component actions *eyebrow oblique*, *lip corner raise* and *eyebrow frown*, respectively. This experiment shows that by using this mapping between emotional states and facial component actions, people infer the same emotional state in the virtual actor as the one used to generate the animations. It also shows that the use of these facial component actions can communicate emotional messages during the speech.

In the Emotionally Expressive Facial Animation System, the selection of a facial signal corresponding a communicative function is carried out according to the emotional state of the virtual actor. This feature enables the EE-FAS to display multiple facial signals that fulfil a communicative function. For instance, a word could be emphasised using different facial signals, such as *eyebrow frown* or *eyebrow oblique.* According to the facial signal selected people could also infer some information about the emotional state of the virtual actor or about the feelings of the virtual actor towards the emphasised word.

6.6 Conclusion

This Chapter presents an experiment based on videos produced by the Emotional Expressive Facial Animation System. This experiment aims to explore the relations between facial component actions and the meanings perceived by people. The advantage of using the EE-FAS to produce the videos for the experiment is the control of which facial component actions are displayed.

By reproducing existing experimental results about perceived emotional states from the displayed facial component actions, this experiment proves that the videos generated by the EE-FAS can communicate the same meanings as those communicated through human faces using identical facial component actions. The experiment also provides some insight about the meanings communicated by the combinations of facial component actions, as well as the meanings communicated by individual facial component action. The influences of individual facial component action and combinations of facial component actions are observed in several emotional dimensions and also on non-emotional dimensions, such as Unfriendly/Friendly and Insincere/Sincere.

The main conclusions drawn from this experiment are that facial component actions influence people's perception of a virtual actor, on emotional and non-emotional dimensions. The perceived meaning of a combination of two facial component actions is generally a combination of the meaning of the individual facial component actions. However, the combination of a smile and an eyebrow frown might influence people's perception of the virtual actor's sincerity differently to the individual facial component action. This has not been statistically proven with the current size sample. This hypothesis is also supported by examples of animated and cartoon characters and more research is planned to examine this situation more closely.

This experiment suggests that the choice of facial signals corresponding to communicative functions is important because they communicate information about the virtual actor, as well as fulfilling their communicative functions. For instance, the facial signal selected to emphasise a word might communicate some information about the emotional state of the virtual actor, or it might communicate some information about what the character feels about the emphasised word.

Chapter 7

Conclusion and Future Work

Virtual actors need a *Dynamic Emotion Representation* to improve their communicative power. This statement is applied to the domain of facial animation, in particular to visual speech.

7.1 Conclusion and Contributions

The aim of this research was to create a facial animation system producing visual speech that can be used by non-professional animators or computer programs. To achieve this, meanings of facial expressions are used as meta-information added to the text to be spoken.

The first issue met was to enable the facial animation system to produce a large number of facial expressions from a small number of facial meanings. From a review of work on human facial expressions and theories of emotion, in Chapter 2, this thesis distinguishes two processes for generating facial expressions: reactions to emotional events and displays of communicative acts. Changes of emotional state produce emotional facial expressions. Communicative acts can display emotional facial expressions, but this thesis also argues that emotional states affect the visual representations of all communicative acts. To extend the number of expressions displayed from facial meanings, this thesis proposes integrating an emotion model within a facial animation system.

From the review of work on synthetic emotions in Chapter 3, computational emotion models are divided into two components: mechanisms eliciting emotions and emotion representations. To control the expressions of a virtual actor it is more important to know its current emotional state than to know what produces this emotional state. Therefore this thesis focuses on emotion representation and emotion dynamics.

Two systems have been developed: the Dynamic Emotion Representation (DER) model and the Emotionally Expressive Facial Animation System (EE-FAS). The EE-FAS integrates an instance of the Dynamic Emotion Representation model in this architecture, to improve the communicative power of the virtual actor.

Three claims are made:

- implementation of a new Dynamic Emotion Representation model enabling programmers to describe a network of persisting states. In this network, the responses of a state to emotional impulses can be influenced by the other states represented in the DER;
- creation of a Dynamic Emotion Representation based on emotion theories and representing three types of state, changing on different timescales;
- use of the Dynamic Emotion Representation to increase the number of facial expressions displayed by the EE-FAS.

7.1.1 The Dynamic Emotion Representation Model

It is recognised that emotions are persisting states but this characteristic of emotions does not get much attention, as the review of emotion theories shows in Chapter 2. The same imbalance is found in computational emotion models reviewed in Chapter 3. The relations and interactions between emotions suffer from the same lack of attention. Most of the research is focused on mechanisms eliciting emotions.

The Dynamic Emotion Representation model presented in Chapter 4 addresses the problem of representing persisting states and the interactions between them.

In the DER model:

- any number of persisting states can be represented;
- each state can be given its own momentum;
- any emotional impulse can affect positively or negatively any state;
- state responses to emotional impulses are influenced by the other states;
- any dynamic emotion representation can be built using an XML file.

7.1.2 A DER Representing Three Momentums

An instance of the DER model has been built to represent three types of state having different momentums: behaviour activations, emotions and moods. This representation is novel because it is based on three timescales where other emotion representation describe a maximum of two. This model can be seen as a mechanism to bias action selection, cognition and expressions of a virtual actor with different degrees of emergency. Behaviour activations represent states that should be taken into consideration immediately; emotions represent contexts in which present matters should be assessed; moods represent contexts in which all matters should be assessed. Moods have less influence than other states.

In Chapter 4 it is shown that the state of this DER depends on the history of previous events and has a momentum changing the effects of future events. It is this DER which has been integrated in the EE-FAS and each type of state has a particular function in the EE-FAS. The behaviour activations are associated with full emotional expressions; the emotions are used to select facial expressions which are speech related; and moods influence the responses of the DER to emotional impulses.

7.1.3 Increasing the Number of Facial Expressions

Due to the presence of the DER in the EE-FAS's architecture, the EE-FAS can display a wider range of facial expressions. The EE-FAS extends the number of facial expressions by:

- distinguishing and producing facial expressions due to emotional events and facial expressions due to communicative acts. These two types of facial expression are different due to their visual aspect, as well as due to their generation processes, as discussed in Chapter 2. Emotional expressions are described as full-face configurations due to sudden emotional events and asynchronous to the speech. Facial expressions related to communicative acts are synchronised with the speech. Chapter 5 presents the EE-FAS's architecture and the role of the DER in the production and selection of these two types of facial expressions.

- displaying different facial signals fulfilling one communicative function according to the DER state. In Chapters 2 and 3 it has been suggested that functions of communicative acts could be fulfilled by different facial signals according to the physical and mental state of the virtual actor and the state of the dialogue. As shown in Chapter 5, the EE-FAS has been implemented to use the DER state as a context to select facial signals corresponding to communicative functions. In Chapter 6 it has been shown that the EE-FAS can fulfil the function of the communicative act *personal reaction*, expressing the emotional state of the virtual actor, by displaying different facial signals according to the DER state. These facial signals are partial-face configurations and yet they are recognised by the subjects as communicating particular emotional messages.

- distinguishing context for displaying genuine and fake facial expressions. The EE-FAS can produce different facial expressions according to whether or not the emotional state of the virtual actor and the intended communicative act are matching. If these two pieces of information do not match, the virtual actor can be seen as lying and in this case the EE-FAS displays a fake facial expressions. This mechanism is described in Chapter 5. The fake expressions can be composed of a masking display which can also be combined with the expression of the felt emotion.

7.1.4 Benefits of the Developed Systems

The benefits of the implementation of the DER model are:

- to maintain any persisting states changing on different timescales;
- to maintain states representing an history of events;
- to enable programmers to develop their own DER or to adapt the DER built from this thesis to their needs, through the configuration of an XML file.

The benefits of the implementation of the EE-FAS are:

- to produce of a wide range of animations from the same script, according to the emotional history of the virtual actor.
- to have a configurable animation system to experiment with the relationships between emotions, communicative acts and facial expressions. These relations are defined in customisable XML files.
- to produce coherent emotional expressions due to the momentum of the DER. In Chapter 5, it is shown that the production of emotional expressions by an emotional impulse is dependent on the state of the DER.
- to produce consistent facial expressions during the speech due to the momentums represented by the DER. The EE-FAS selects facial signals according to the state of the DER; therefore over a certain period of time identical facial signals would be selected to express a communicative act. In addition, if the communicative functions are well defined, the number of conflicts between facial signals should be limited because non-conflicting facial signals can be defined for each emotional context.

7.2 Secondary Contributions

In Chapter 5 a new categorisation of facial expression is suggested. This categorisation merges Ekman's description of facial expressions (Ekman, 1979) and

Chovil and Bevelas's list of communicative functions (Bavelas and Chovil, 1997). It also distinguishes four types of facial expression related to emotions: *emotional expressions* which are reactions to emotional events and based on the universally recognised facial expressions; *portrayal* which can be symbolic, exaggerated and culturally dependent emotional expressions; and *personal reactions* which are displays of believable emotional expressions divided into expressions of *felt* and *non-felt* emotions. The latter represents the distinction between a genuine and fake smile.

The experiment described in Chapter 6 shows that people can recognise emotional messages through the interpretation of certain movements of facial parts, in contrast to full-face configurations, such as the universally recognised facial expressions. It has also shown that the meanings communicated by combinations of facial component actions are most of the time combinations of the meanings of individual facial component actions. One exception could be the combinations of a smile and an eyebrow frown, which communicates insincerity where individual facial component actions do not.

7.3 Future Work

The Dynamic Emotion Representation described in this dissertation is used to represent states changing on different time scales. During the creation of videos for the experiment or for the examples, the DER has been used only over a few minutes. An extention of this work could be to experiment with longer term situations, over hours, days or weeks and see how the DER helps to produce believable behaviours. One method would be to integrate the DER in an autonomous agent architecture and measure the differences in behaviours with and without the DER.

The DER model has been used to represent emotional states such as happiness and anger. These states might not be directly represented in natural systems; it is difficult to believe that somewhere inside people a "little box" keeps the level of anger. Instead, the state of anger might be better represented by levels of certain chemicals in the body. The DER presented in this thesis is a high level representation of particular states which can be used to influence other mechanisms without having to represent all chemical levels of the body. To create autonomous virtual actors behaving in a human-like fashion, the level of detail of the representations of internal states needed should be explored. The emotional categorisations described in this thesis are intuitive representations to control virtual actors but lower level descriptions of internal states might be needed to create autonomous virtual actors. For instance, the design of an architecture for virtual autonomous actors will be successful when millions of people will watch "virtual reality shows" on television played by characters based on this architecture.

Part IV

Appendix

Appendix A

Meanings Communicated by Facial Components

Table A.1: Pleasure and Arousal scores for single Action Units

AU	Description	Pleasure	Arousal
1	Inner brow raised	-1.92	-0.85
2	Outer brow raised	1.85	2.11
4	Brow furrowed	-2.00	1.23
5	Upper eyelid raised	-1.19	3.04
6	Cheek raised	1.46	0.73
7	Lower eyelid raised	-1.85	0.58
9	Nose wrinkled	-2.33	2.33
10	Upper lip raised	-3.67	1.97
15	Lip corner depressed	-2.26	-0.63
17	Chin raised	-0.93	0.60
46	Wink	1.19	0.54

Table A.2: Important components of some widely recognized facial expressions (Smith and Scott, 1997, Table 10.1; p. 233)

	Facial action						
	Eyebrow frown	Raise eyebrows	Raise upper eyelid	Raise lower eyelid	Lip corners	Open mouth	Raise upper lip
Muscular basis	currugator supercilii	medial frontalis	levator palpebrae superioris	orbicularis oculi	a	orbicularis oris	levator labii superioris
Action Units	4	1	5	6,7		26,27	9,10
Emotion expressed							
Happiness				X	Raise	X	
Surprise		X	X			X	
Anger	X		X	X			
Disgust/				X			X
contempt	X						
Fear	X	X	X			X	
Sadness	X	X			Lower		

Table A.3: Meanings proposed to be associated with individual facial actions (Smith and Scott, 1997, Table 10.2; p. 238)

	Facial action							
Meaning	Eyebrow frown	Raise eyebrows	Raise upper eyelid	Raise lower eyelid	Lip corners	Open mouth	Tighten mouth	Raise chin
Pleasantness	−				+	+	−	−
Goal obstacle/ goal discrepancy	+							
Anticipated effort	+							
Attentional activity		+	+			+		
Certainty		−		+				
Novelty		+	+					
Personal agency/ control		−	−			−		

Appendix B

Formalism to Combine Facial Signals

B.1 Introduction

As introduced in Chapter 5 facial meanings are transformed into facial signals which are the physical representations of meanings.

When multiple meanings need to be communicated at the same time, their corresponding facial signals need to combine. Facial signals are composed of Facial Component Actions (FCAs), which are movements of facial parts. Two facial signals can be combined in many ways, the result is another facial signal composed of some of the FCAs from both facial signals. Rules and priorities need to be explicit to be able to control the combination process. This Chapter presents a formalism describing facial signals, operators to manipulate facial signals and their components, and a set of equations used to combine two facial signals. The latter also describes a priority mechanism between two facial signals, transferring the influence on the result from one facial signal to the other.

This formalism provides a set of attributes to the user, describing characteristics of a facial signal and its components, to control the combination process. These attributes relate to how the meaning of a facial signal is carried. As described in Section 2.2, certain facial signals need all their FCAs to be displayed to communicate their meanings and others only need certain of their components to be displayed to communicate their meanings. This type of information is used in the combination process to produce a facial signal that communicates as well as possible the meanings of the two combined facial signals.

One issue is the occurrence of semantic and physical conflicts during these combinations. A semantic conflict occurs when two facial signals communicating two different meanings with the same facial components are combined. Depending on the type of combination used, the result might communicate a meaning which is not intended. The formalism described in this Chapter enables users to describe the characteristics of facial signals and to eliminate the semantic conflicts. Physical conflicts happen when the combination of two facial signals produces facial movements which are not physically possible. These conflicts are taken into consideration and simple techniques are described to solve this issue.

This Chapter is based on the Emotionally Expressive Facial Animation System (EE-FAS) as a practical ground to develop and test the formalism. Throughout this Chapter the definitions will evolve to adapt to the problems met. Examples are shown and discussed, making the issues and solutions clearer. This Chapter refers several time to Levels of Control; these Levels of Control (LoC) are defined in Chapter 1.

After this introduction, the Chapter carries on with Section B.2 issues of combining facial signals. Section B.3 describes the existing work of others related to the combination of facial expressions and Section B.4 presents a formalism to combine facial signals. Section B.5 touches on the problems of animation discontinuities due to the combination process and proposes some solutions to resolve them. Section B.6 presents some results of different combinations with figures showing how the components of facial signals are combined and the resulting screen-shots.

B.2 Issues of Combining Facial Signals

This Chapter uses the Emotionally Expressive Facial Animation System (EE-FAS) as a practical ground to explain and test the issues occurring during the combinations of facial signals.

B.2.1 Practical Example with the EE-FAS

The EE-FAS plays facial animation scripts composed of tags representing communicative functions, emotional impulses and visemes. Emotional impulses can trigger emotional behaviours, such as Happy or Surprised, which are communicated through emotional facial expressions, such as a smile. The emotional behaviour is the meaning that should be communicated through a facial signal which is the physical movement of the face. Communicative functions also represent facial meanings that should be communicated through facial signals.

When animation is played, the transformations from facial meanings into facial signals are carried out in real-time. EE-FAS is based on a modular architecture and different modules are used to transform communicative functions and emotional behaviours into facial signals. One module receives all the facial signals. This module, called Facial Component Action Dynamic Representation (FCADR), keeps track of the facial movements described by facial signals and combines facial signals together. In fact, facial meanings can overlap. In this case several facial signals can be displayed at the same time, which is the reason for combining facial signals. Every so often, the FCADR module takes a snapshot of the current configuration of the face to be displayed on the screen, producing an animation.

B.2.2 Issues in Combining Facial Signals

A facial signal is the physical representation of a meaning and by the fact represents a meaningful unit. Facial signals are composed of Facial Component Actions (FCAs), which describe actions occurring in parts of the face, e.g. in a facial component. Associated with each FCA are functions of time describing the changes of action intensities, describing movements. An FCA can be *right eyebrow frown* or *right lip corner raised.* Three functions of time are used to describe the intensity changes, one for the attack duration, one for the sustain duration and one for the decay duration of the Facial Component Action.

During semantic conflicts, the combination of the Facial Component Actions can produce a meaning which is different from the ones that should be communicated. For instance, the desire to express an order can be communicated through an eyebrow frown action and at the same time a word can be emphasised with an eyebrow raise action. The combination of these two movements communicates a meaning of sadness, which is not the intended meaning (Pelachaud and Bilvi, 2003). Solutions for both types of conflicts are implemented in the EE-FAS.

A first issue is to define a method to combine facial signals, to define a mechanism that selects which parts of which facial signal should be selected to communicate as well as possible the meanings of both facial signals.

A second problem is to define a priority mechanism to select which parts of the conflicting facial signals should be kept in the facial signal resulting from the combination.

A particular issue also occur when a facial signal displayed on the screen is involved in a conflict. If parts of this facial signal were stopped due to conflicts a facial movement would be interrupted and jumps in the screen animation would appear. We call this problem animation discontinuity.

B.2.3 Foundations for a Combination Formalism

As introduced earlier, facial signals are physical implementation of meanings, but people's views on how meanings are carried out by facial signals differ. The first approach is the *purely categorical approach*, suggesting that facial meanings are communicated only through full-face configurations. The parts of these configurations, e.g. Facial Component Action, do not communicate any meanings by themselves (Smith and Scott, 1997). The second approach is the *componential approach*, in which movements of facial part have meanings and the meaning of the full-face configuration is a combination of the meanings of its parts (Smith and Scott, 1997). This distinction is important when facial signals are combined. In the case of the purely categorical approach, facial signals cannot be combined without losing their meanings, whereas with the componential approach the facial signals can be combined. Both views can be used in the EE-FAS and the mechanism for combining facial signals is based on both type of descriptions.

To design a combination process which keeps meanings of facial signals while reducing conflicts, a formal definition of facial signals has been created. This formal definition describes a facial signal and the relation between the meanings communicated by the facial signal and the facial signal itself. To select and to combine parts of facial signals operators are defined and using these operators a set of equations are designed to describe several ways to combine two facial signals. When conflicts occur during a combination a priority mechanism is needed to select one of the conflicting elements. This priority mechanism also is defined by the set of equations.

B.3 Related Work

This section presents the main solutions to combine facial expressions suggested by other researchers, as well as methods to reduce animation discontinuities due to semantic or physical conflicts.

B.3.1 Animation Parameters: Combination by Addition

Animation parameters are commands at Level of Control 2 and enable animators to modify graphical representations, which are at Level of Control 1, using a small set of values. For instance, the change of the contraction value of an Abstract Muscle modifies the position of all the vertices within its zone of action (Parke and Waters, 1996). Several types of animation parameters exist: Abstract Muscles, Action Units defined by the Facial Action Coding System, and Facial Animation Parameters defined by MPEG-4 (Parke and Waters, 1996; The Duy Bui, 2004; Ekman and Friesen, 1978; Pirker and Krenn, 2002).

Facial expressions can be described by a small set of facial animation parameters and the combination of two facial expressions can be carried out by adding together values of the parameter (DeCarlos et al., 2002; Albrecht et al., 2002; Pelachaud and Bilvi, 2003). The problem with this method is that the value of an animation parameter can be excessive and so produce unrealistic results.

The combination of actions of certain animation parameters can also produce unrealistic results (Bui et al., 2004).

B.3.2 Algebra of Expression

Paradiso defines an Algebra of Expression (AoE) to combine facial expressions. A facial expression is defined as a vector of 66 Facial Animation Parameter intensities, Facial Animation Parameters numbered from 3 to 66 by MPEG-4. This description is static and therefore represents command at Level of Control 2.

A facial expression S is represented by $S(s_1, \ldots, s_{66})$ where s_i represent a Facial Animation Parameter intensity, as an integer. To manipulate this representation, three operators are defined: *sum*, *amplifier* and *overlapper*.

Suppose $S = (s_1, \ldots, s_{66})$ and $T = (t_1, \ldots, t_{66})$ are two facial states.

The operator *sum* is defined as
$S +_v T = |s_i.v| + |t_i.v|$ where the symbol $||$ denotes the approximation to the closest integer and $v \in [0, 1]$.

The operation *amplifier* is a particular case of the *sum* operator.

The operator *overlapper* is defined as

$$(g_i) = \begin{cases} g_i = s_i + t_i & \text{if } s_i \text{ or } t_i = 0 \\ g_i = s_i & \text{if } p_1 > p_2 \\ g_i = t_i & \text{otherwise} \end{cases}$$

Where $i = 1, \ldots, 66$ and p_1 and p_2 are the priorities of S and T respectively.

The definitions of these operators are extended to manipulate animations. An animation is defined as a sequence of facial states, which means that animations are described by discrete time steps, in a key frame fashion. The operators act in the same way as previously described when the time steps of two animations overlap in time.

These operators act at Level of Control 2 because they manipulate static facial representations. This algebra of expression does not resolve any types of conflict.

B.3.3 MIMIC: an Animation Programming Language

MIMIC can be described as an animation programming language enabling animators to specify sequences of groups of facial actions (Fuchs et al., 2004). This language is independent of the type of animation parameters used to describe the face. Movements of the face can be defined by several types of function of time, such as exponential or linear function, describing how animation parameter values change over time. This type of description is at Level of Control 3 because it adds functions of time to the descriptions at Level of Control 2.

MIMIC defines actions as an association, or a group of associations between a time function and an animation parameter. The compiler for this language resolves the synchronisation constraints between the actions defined by the user and blends actions on a same animation parameter in several customisable fashions, such as by computing the average or the sum of the action intensities.

MIMIC does not resolve any types of conflict. A difference between a system using MIMIC and EE-FAS is that MIMIC does not work in real-time, it has access to the complete animation to resolve synchronisation issues and uses knowledge about next facial expressions to do the blending between two actions.

B.3.4 Combination of Facial Movements

In the system described by Bui et al. (2004), facial movements are classified by channels, such as *manipulator*, *lip movement* and *emotion display* channels. In each channel two facial movements cannot occur at the same time but their ending and beginning can overlap. Several channels can be active at the same time, combining the facial movements.

A facial movement is described over three periods of time: attack, sustain and decay. In addition, the facial movement is also described by a list of intensities reach by each animation parameter after the attack period.

This paper describes an algorithm to blend transitions between two facial movements affecting the same animation parameter. The blend algorithm changes the starting intensity of the animation parameter from the latest facial movement to the current intensity of this animation parameter.

This paper also approaches the issue of physical conflicts and suggests detecting them by using a list of animation parameters that cannot be active at the same time. In this system, physical conflicts are detected before they occur, and the facial movement affecting the animation parameter on which the conflict will occur is slowly interrupted before the second conflicting facial movement starts.

This system does solve some semantic conflicts through its use of different channels for different communicative acts. The problem with this solution is that movements from different channels are combined together, which can produce semantic conflicts. For instance, there cold be two facial movements active in two different channels, such as a *conversational signal* of *emphasise*, represented by eyebrow raise, and an *emotion display* of *anger*, represented by eyebrow frown. The combination of these two facial movements would produce an expression of sadness, which is not the meaning intended.

The solution to resolve conflicts, proposed by this paper, cannot be applied in the EE-FAS because it takes advantage of knowing the next facial movement. Also, this system does not solve all semantic conflicts.

B.3.5 One Meaning per Facial Parts

The system described by Pelachaud and Bilvi (2003) implements a mechanism to combine facial expressions to create new ones and also a solution to resolve semantic conflicts. This system defined facial parts where meanings are expressed. Only one meaning can be expressed at one time in each facial parts. The selection mechanism is implemented using a Belief Network.

A facial expression is defined as a vector of 66 Facial Animation Parameters and the combination of two facial expressions is carried out by summing the values of corresponding Facial Animation Parameters.

B.4 Formalism to Combine Facial Signals

Facial animation systems need to combine facial signals or facial expressions, to express multiple meanings at the same time or to produce a smooth transition between two facial signals. This section gives a formal description of a facial signal, including functions of time and the relationship between facial signals and meanings. Operators are defined to manipulate facial signals and to create a set of equations used to combine facial signals and describe a priority mechanism. This formalism is based on the fact that meanings of facial signals need to be kept when they are combined. This formalism is not only designed to explain how facial signals can be combined but is suitable to be implemented in the EE-FAS. This means that all the features needed to combine and to represent facial movements, e.g. facial signal, have to be in the definition.

To help to understand the formalism, this section explains how this formalism has been created and extends the definitions along the way to adapt them to the issues met.

B.4.1 First Facial Signal Definition

The face is divided into components called *Facial Components*, such as *right eyebrow* or *right lip corner*. Each Facial Component has a list of possible actions, called *Facial Component Actions.* A Facial Component Action can be *right eyebrow frown* or *left lip corner raised.* More examples can be found in Table B.1.

A facial signal can be defined as pairs of a facial component and a facial component action, such as $FS = \{(C_l, A_{l,m})\}$ where C_l is a facial component, $A_{l,m}$ a facial component action and l and m are indices as shown in Table B.1. Because it is possible to find a facial component C_l from a facial component action $A_{l,m}$ it should not be necessary to have C_l in the representation. The reason for including C_l in the definition is to be able to manipulate the facial signal at a facial component level or at a facial component action level. This will become clearer later in this section.

This type of representation is static and facial signals are facial movements: the

Table B.1: Examples of Facial Components and Facial Component Actions

	Facial Component	Right Eyebrow	Right Upper Eyelid	mouth
Indices	$m \setminus l$	1	2	3
Facial Component Action	1	right_eyebrow _raised	right_upper _eyelid_open	right_lip _corner_raised
	2	right_eyebrow _frown	right_upper _eyelid_close	right_lip _corner_lower
	3	right_eyebrow _oblique		viseme_O

intensities of actions change over time. Therefore a facial signal is defined as a list of three elements: a Facial Component (C_l), a Facial Component Action ($A_{l,m}$) and a function of time describing the changes of FCA intensity (T_i).

Definition of Facial Component 1. *A Facial Components C_l is member of the set of all the facial components. Facial components are part of the face, such as* right eyebrow *or* right lip corner. *Two facial components C_l and $C_{l'}$ are the same if and only $l = l'$.*

Definition of Facial Component Action 1. *A Facial Component Action $A_{l,m}$ is a member of a set of facial component actions indexed by facial components, as described in Table B.1. To each C_l is associated a set of $A_{l,m}$. For instance, a facial component action $A_{l,m}$ could be* right eyebrow raised. *Two facial component actions $A_{l,m}$ and $A_{l',m'}$ are the same if $l = l'$ and $m = m'$.*

Definition of Facial Signal 1. *A Facial Signal FS is a set of triplets: a facial component C_l, a facial component action $A_{l,m}$, and a function of time returning the facial component action intensity T_i. The time function is made of three phases: attack, sustain and decay (DeCarlos et al., 2002; Kalra et al., 1991; The Duy Bui, 2004; Pelachaud and Bilvi, 2003).*

$FS : \{(C_l, A_{l,m}, T_i)_{l,m,i}\}$

where:

- $(C_l, A_{l,m}, X)_{l,m}$ *is unique,* X *being any function of time. This is to limit conflicting controls of the same group of animation parameters.*

- $C_l \in FC$. *FC is the set of Facial Components, where each element is unique.*

- $A_{l,m} \in FCA_l$. *FCA_l is the set of the Facial Component Actions affecting one Facial Component C_l, where each element is unique. Each $A_{l,m}$ is related to one particular C_l.*

- $C_l = C_{l'}$ *if* $l = l'$.

- $A_{l,m} = A_{l',m'}$ *if* $l = l'$ *and* $m = m'$.

- $T_i = (is, ie, stime, \{(F_n, st_n, et_n)_n\})$. *$T_i$ can be seen as a "function of time" returning the intensity of $A_{l,m}$. is, ie and stime are respectively $A_{l,m}$'s intensity at the start, $A_{l,m}$'s intensity at the end and the time at which the $A_{l,m}$ starts. $\{(F_n, st_n, et_n)_n\}$ is a set of functions of time describing a continuous function. st_n is the time at which the function F_n starts and et_n is the time at which the function F_n ends.*

  Typically, F_n contains three functions describing the changes of intensity during the attack, the sustain and the decay durations.

B.4.2 Interesting Operators

To combine Facial Signals, which are sets of elements, operators need to be defined and these operators must generate interesting sets. What are the interesting set operations to combine two sets?

One interesting operation is the *Difference* between two sets. The difference operator provides all the elements which are in the set FS1 but not in the set FS2, where FS1 and FS2 are two facial signals. This operator is *not commutative.* This operator enables us to separate elements which are unique to FS1.

Another interesting operation is the *Intersection* of two sets. This operation produces a set with all the elements which appear in both sets.

Finally, it is useful to produce a set from two sets and this operation could be the *Union* operation. The union operation produces a set with all the elements from both sets.

These three operations help to build new facial signals from two original facial signals, but they all assume that it is easy to compare one element to another, e.g. to decide if an element is part of a set or not.

With the facial signal definition 1, the elements composing a facial signal are made of three sub-elements: C_l, $A_{l,m}$ and T_i. Therefore what does mean so say that two elements are identical or different?

An element is the same to another one if all three sub-elements are the same. A function of time T_i describes the changes of intensity of a particular facial component action, but it is not interesting to compare elements on this basis because too many features of T_i can vary, such as start and end time. These values are necessary to describe completely a movement but they are not significative to knowing which parts of two facial signals should be combined. Instead of comparing elements of facial signals using temporal information, it seems more reliable to compare them on physical features, such as Facial Components and Facial Component Actions.

Another solution is to reduce the comparison of elements to the comparison of a pair of the sub-elements $(C_l, A_{l,m})$. A particularity of these pairs is that for each sub-element $A_{l,m}$ the other sub-element C_l is fixed because each Facial Component Action is associated with one Facial Component. The comparison of elements using this pair of sub-elements is not a good choice because the same result can be obtained by comparing only the sub-elements $A_{l,m}$.

It is possible to compare elements using either the sub-element C_l or the sub-element $A_{l,m}$, which gives different results. Therefore, the three operators need to be indexed by the type of comparison used, either C or A: $Difference_C$, $Intersection_C$, $Union_C$, $Difference_A$, $Intersection_A$ and $Union_A$.

What is the meaning of using the operators indexed by C or by A?

For instance let's define two Facial Signals FS_1 and FS_2:

$$FS_1 : \{(C_1, A_{1,1}, T_1), (C_1, A_{1,2}, T_2), (C_2, A_{2,1}, T_3)\}$$
$$FS_2 : \{(C_1, A_{1,1}, T_4), (C_1, A_{1,3}, T_5), (C_3, A_{3,1}, T_6)\}$$

and two operators *Difference* noted as $\backslash_C$ and $\backslash_A$. Note that these operators are not commutative.

$$FS_1 \backslash_C FS_2 = \{(C_2, A_{2,1}, T_3)\}$$
$$FS_1 \backslash_A FS_2 = \{(C_1, A_{1,2}, T_2), (C_2, A_{2,1}, T_3)\}$$

The operator $\backslash_C$ looks for facial component C_l only present in FS_1. The operator $\backslash_A$ looks for facial component action $A_{l,m}$ only present in FS_1.

The operator $\backslash_C$ is useful to discover which facial component is free of any action. This method could be used to find a free facial component that could be used to express a new meaning. This operator could be used to create a feedback loop of the face.

What happens with the operators *Intersection*: $\cap_{fc}$ and $\cap_{fca}$?

$$FS_1 \cap_C FS_2 = \left\{ \begin{array}{c} (C_1, A_{1,1}, T_1), (C_1, A_{1,2}, T_2), \\ (C_1, A_{1,1}, T_4), (C_1, A_{1,3}, T_5) \end{array} \right\}$$

$$FS_1 \cap_A FS_2 = \left\{ \ (C_1, A_{1,1}, T_1), (C_1, A_{1,1}, T_4) \ \right\}$$

The same problem occurs for these two operators because T_1 and T_4 are different functions of time, with different end times, start times or shapes. Using the index C or A, the element $(C_1, A_{1,1}, X)$ is not unique and the issue is to deal with a command driven by two different functions of time T_1 and T_4 controlling the same group of animation parameters.

The intersection of two facial signals is very important in the combination process because this is where the problems occur.

It would be better to define some rules to control which elements should be part of the intersection between two sets. In a new definition of a facial signal, a weight is associated to each triple $(C_l, A_{l,m)}, T_i}$. These weights are used to select facial component actions in the operation $Intersection_C$ and $Intersection_A$.

The same problem appear with the operator *Union* because it includes the intersection of two sets, therefore instead of this operator the formalism defined the operators $XUnion_C$ and $XUnion_A$. The operator $XUnion$ produces a set of all the elements which are only in one of the two sets. These operators are commutative.

$$FS_1 \cup_C^X FS_2 = \left\{ \ (C_2, A_{2,1}, T_3), (C_3, A_{3,1}, T_6) \ \right\}$$

$$FS_1 \cup_A^X FS_2 = \left\{ \begin{array}{c} (C_1, A_{1,2}, T_2), (C_2, A_{2,1}, T_3) \\ (C_1, A_{1,3}, T_5), (C_3, A_{3,1}, T_6) \end{array} \right\}$$

B.4.3 First Operator Definitions

The temporary new definition of a facial signal is:
$FS : \{((C_l, A_{l,m}, T_i), W_e)_{l,m,i,e}\}$ where W_e is the weight of $(C_l, A_{l,m}, T_i)$.

Let's define two Facial Signals:
$FS_1 : \{((C_l, A_{l,m}, T_i), W_e)_{l,m,i,e}\}$ and $FS_2 : \{((C_k, A_{k,n}, T_j), W_f)_{k,n,j,f}\}$

In Operator 1. $((C_l, A_{l,m}, T_i), W_e) \in FS_1$.

$((C_l, A_{l,m}, T_i), W_e) \in FS_1$ *if* FS_1 *contains* $((C_l, A_{l,m}, T_i), E_e)$.

In Operator 2. $((C_l, A_{l,m}, T_i), W_e) \in_C FS_1$.

$((C_l, A_{l,m}, T_i), W_e) \in_C FS_1$ *if an element of* FS_1 *contains the sub-element* C_l.

In Operator 3. $((C_l, A_{l,m}, T_i), W_e) \in_A FS_1$.

$((C_l, A_{l,m}, T_i), W_e) \in_A FS_1$ *if an element of* FS_1 *contains the sub-element* $A_{l,m}$.

Not In Operator 1. $((C_l, A_{l,m}, T_i), W_e) \notin FS_1$.

$((C_l, A_{l,m}, T_i), W_e) \notin FS_1$ *if* FS_1 *does not contain* $((C_l, A_{l,m}, T_i), W_e)$.

Not In Operator 2. $((C_l, A_{l,m}, T_i), W_e) \notin_C FS_1$.

$((C_l, A_{l,m}, T_i), W_e) \notin_C FS_1$ *if* C_l *is not a sub-element of any element of* FS_1.

Not In **Operator 3.** $((C_l, A_{l,m}, T_i), W_e) \notin_A FS_1$.

$((C_l, A_{l,m}, T_i), W_e) \notin_A FS_1$ if $A_{l,m}$ is not a sub-element of any element of FS_1.

Difference **Operator 1.** $FS_1 \setminus FS_2$.

The $Difference$ operator builds $\{((C_v, A_{v,x}, T_y), W_z)_{v,x,y,z}\}$ where $((C_v, A_{v,x}, T_y), W_z) \in FS_1$ AND $((C_v, A_{v,x}, T_y), W_z) \notin FS_2$.

This operator is not commutative.

Difference **Operator 2.** $FS_1 \setminus_C FS_2$.

The $Difference_C$ operator builds $\{((C_v, A_{v,x}, T_y), W_z)_{v,x,y,z}\}$ where $((C_v, A_{v,x}, T_y), W_z) \in_C FS_1$ AND $((C_v, A_{v,x}, T_y), W_z) \notin_C FS_2$.

This operator is not commutative.

Difference **Operator 3.** $FS_1 \setminus_A FS_2$.

The $Difference_A$ operator builds $\{((C_v, A_{v,x}, T_y), W_z)_{v,x,y,z}\}$ where $((C_v, A_{v,x}, T_y), W_z) \in_A FS_1$ AND $((C_v, A_{v,x}, T_y), W_z) \notin_A FS_2$.

This operator is not commutative.

XUnion **Operator 1.** $FS_1 \cup^X FS_2$.

The XUnion *operator builds $\{((C_v, A_{v,x}, T_y), W_z)_{v,x,y,z}\}$ where $((C_v, A_{v,x}, T_y), W_z) \in FS_1$ XOR $((C_v, A_{v,x}, T_y), W_z) \in FS_2$.*

XUnion **Operator 2.** $FS_1 \cup^X_C FS_2$.

The XUnion *operator builds $\{((C_v, A_{v,x}, T_y), W_z)_{v,x,y,z}\}$ where $((C_v, A_{v,x}, T_y), W_z) \in_C FS_1$ XOR $((C_v, A_{v,x}, T_y), W_z) \in_C FS_2$.*

This operator is commutative.

XUnion **Operator 3.** $FS_1 \cup^X_A FS_2$.

The XUnion *operator builds $\{((C_v, A_{v,x}, T_y), W_z)_{v,x,y,z}\}$ where $((C_v, A_{v,x}, T_y), W_z) \in_A FS_1$ XOR $((C_v, A_{v,x}, T_y), W_z) \in_A FS_2$.*

This operator is commutative.

Intersection Operator 1. $FS_1 \cap_C FS_2$

The Intersection$_C$ operator builds $\{((C_v, A_{v,x}, T_y), W_z)_{v,x,y,z}\}$
where $((C_v, A_{v,x}, T_y), W_z) \in_C FS_1$ *AND* $((C_v, A_{v,x}, T_y), W_z) \in_C FS_2$.

When $v = l = p$:

If $w_l \geq u_p$ *then* $x = m$, $y = i$ *and* $z = e$.

If $w_l < u_p$ *then* $x = q$, $y = j$ *and* $z = f$.

This operator is not commutative.

Intersection Operator 2. $FS_1 \cap_A FS_2$

The Intersection$_A$ operator builds $\{((C_v, A_{v,x}, T_y), W_z)_{v,x,y,z}\}$
where $((C_v, A_{v,x}, T_y), W_z) \in_A FS_1$ *AND* $((C_v, A_{v,x}, T_y), W_z) \in_A FS_2$.

When $v = l = p$ *and* $x = m = n$

If $w_l \geq u_p$ *then* $y = i$ *and* $z = e$.

If $w_l < u_p$ *then* $y = j$ *and* $z = f$.

This operator is not commutative.

B.4.4 Equations to Combine Facial Signals

The previous definitions of facial signals and operators are not complete but it is now possible to set up the basis for a set of equation that could be used to combine facial signals.

The facial signals FS_1 and FS_2 are the two operands in the following equations and FS_r is the resulting facial signal.

For simplicity, the operators $\cup^X$, $\setminus$ and $\cap$ are used instead of respectively the operators $\cup_C^X$ or $\cup_A^X$, $\setminus_C$ or $\setminus_A$ and $\cap_C$ or $\cap_A$. It should be noticed that the operators $\cap_C$ and $\cap_A$ are not commutative.

With the operators $\setminus$ and $\cap$ it is possible to define the following sub-sets from two facial signals:

- FS_1: All the elements from FS_1.
- FS_2: All the elements from FS_2.
- $FS_1 \setminus FS_2$: The elements only present in FS_1.
- $FS_2 \setminus FS_1$: The elements only present in FS_2.
- $FS_1 \cap FS_2$: The elements present in both FS_1 and FS_2 but with a preference for FS_1 when choices need to be made.
- $FS_2 \cap FS_1$: The elements present in both FS_1 and FS_2 but with a preference for FS_2 when choices need to be made.

These sub-sets can be "combined" using the operator $\cup^X$. The number of possible combinations is ${}_6C_0 + {}_6C_1 + {}_6C_2 + {}_6C_3 + {}_6C_4 + {}_6C_5 + {}_6C_6 = 64$. It should be noticed that some results of these combinations are not intuitive. For instance, FS_1 is not the same as $(FS_1 \setminus FS_2) \cup^X (FS_1 \cap FS_2)$ due to the use of weights to select elements going into the intersection $FS_1 \cap FS_2$. On the other hand, $FS_1 \cup^X FS_1$ is the same as $FS_1 \setminus FS_2 \cup^X FS_2 \setminus FS_1$.

The problem is the difficulty to use all these combinations in a system because there are not obvious reasons to select one combination more that another. From all these combinations, only 8 of them have been selected to describe how much two facial signals influence the resulting facial signal, to define a priority mechanism between the two facial signals.

Equations for Combining Facial Displays 1.

$$FS_1 = FS_r \tag{B.1}$$
$$FS_1 \cup^X (FS_2 \setminus FS_1) = FS_r \tag{B.2}$$
$$(FS_1 \setminus FS_2) \cup^X (FS_1 \cap FS_2) = FS_r \tag{B.3}$$
$$(FS_1 \setminus FS_2) \cup^X (FS_1 \cap FS_2) \cup^X (FS_2 \setminus FS_1) = FS_r \tag{B.4}$$
$$(FS_2 \setminus FS_1) \cup^X (FS_2 \cap FS_1) \cup^X (FS_1 \setminus FS_2) = FS_r \tag{B.5}$$
$$(FS_2 \setminus FS_1) \cup^X (FS_2 \cap FS_1) = FS_r \tag{B.6}$$
$$FS_2 \cup^X (FS_1 \setminus FS_2) = FS_r \tag{B.7}$$
$$FS_2 = FS_r \tag{B.8}$$

where $\cup^X$ *could be either* $\cup_C^X$ *or* $\cup_A^X$, $\setminus$ *could be either* $\setminus_C$ *or* $\setminus_A$ *and* $\cap$ *could be either* $\cap_C$ *or* $\cap_A$. *In fact the type of operator used is not specified at a facial signal level but on a facial component action basis, and this is explained later in this section.*

Equation C.1 describes a case where the FS_1 has a complete influence on the result, which means that FS_1 has total priority over FS_2. As a logical regression of FS_1's priority, Equation C.2 enables FS_2 to contribute to the resulting facial signal by adding its elements which are not in FS_1 and all the elements of FS_1 are conserved unchanged. Equation C.3 provides even more influence to FS_2, enabling it to play a role by eventually modifying the elements in common with FS_1. Equation C.4 enables FS_2 to play a role based on the common elements with FS_1 but also to add its elements which are not in FS_1 to the resulting facial signal. Equations C.4 and C.5 are different due to the non-commutativity of the operator *Intersection.* Due to this characteristic Equation C.5 provides more influence to FS_2. The rest of the equations can be described in the same way but inverting the priorities between FS_1 and FS_2.

B.4.5 Combinations and Facial Meanings

The aim of this formalism is to combine facial signals with regard to how their meanings are carried and to create a combined facial signal which communicates as well as possible the meanings of the two original facial signals.

Attributes *type* and *priority* to select a Combination

As discussed in Section 2.2.2, a meaning of a facial signal can be carried:

- by the complete facial signal, e.g. all the facial component actions must be shown to express the meaning of this facial signal;
- Selectively by some facial components or some facial component actions of this facial signal.

To reflect these differences, an attribute called *type* is added to the definition of a facial signal, specifying how the meaning is carried. This attribute can have the values *categorical* when the meaning is carried by the complete facial signal or *componential* when the meaning is carried more strongly by some of the facial signal components (Smith and Scott, 1997).

Equations C.1 or C.2 can be used to combine a facial signal FS_1 which has the type *categorical* because they are the only equations ensuring that all the facial component actions of FS_1 will be present in the result of the combination. Facial signals having the type *componential* can be combined with any equation from the set.

To select which equation should be used to combine two facial signals, another attribute is added to the definition of a facial signal. This attribute is called *priority*.

The two attributes *type* and *priority* qualify a facial signal and their values are used to select an equation to combine this facial signal with another. In practice, the EE-FAS displays a current facial signal, which is combined with new facial signals that need to be displayed. The attributes of new facial signals, in contrast with the attributes of the current facial signal, are used to select the equation combining these two facial signals.

So a new temporary definition of a facial signal is:
$FS : (\{((C_l, A_{l,m}, T_i), W_e)_{l,m,i,e}\}, type, priority)$.

Attribute *TC* to Select the Indices of the Operators

A *facial component* could be affected by several *facial component actions*. In certain cases during the combination of two facial signals, it is interesting to limit the number of meanings communicated by one facial component, e.g. to limit the number of facial component actions occurring in a facial component. The limitation to one facial component action per facial component eliminates the occurrence of semantic conflicts. An example of semantic conflict is the display of a frown communicating an order and the display of eyebrows raised, emphasising a word. These actions occur on the eyebrow and the combination of these actions would communicate a meaning of sadness. By limiting the number of actions on one facial component to one, this conflict would not occur, which is the method used by Pelachaud and Bilvi (2003). Some other facial components, such as the mouth, need to be able to have several facial component actions at one time: the combination of visemes and a smile (right_lip_corner_raised and left_lip_corner_raised), for instance.

Therefore a new attribute is added to the definition of a facial signal. This attribute, called *TC* for Type of (facial) Component, is associated each facial component and takes the value *single* if it should have only one facial component action or *multi* if it can have multiple facial component actions.

So a new temporary definition of a facial signal is:
$FS : (\{((C_l, A_{l,m}, T_i), W_e, TC_g)_{l,m,i,e,g}\}, type, priority)$.

Examples of how this attribute is used to select the indices of the operators are shown in the next sub-section.

Attribute *BL* to Blend Identical Facial Component Action

If, in certain cases, two different facial component actions cannot occur at the same time on the same facial component, it is possible to have the same facial component action with two different intensities on the same facial component without modifying the meaning. For instance, in the earlier example, a frown is used to communicate an order but a frown can also be used to emphasise a word, instead of eyebrows raise. In this case, both meanings can be communicated without creating a conflict. Both intensities of this facial component action could be *added*, *averaged*, or one could *overwrite* the other, perhaps depending on the intensities.

To enable the formalism to cope with these situations a final attribute, called *BL* for type of BLend, is added to the definition. This attribute can take one of the following values: *add*, *average* and *overwrite*.

So a new temporary definition of a facial signal is:
$FS : (\{((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a)_{l,m,i,e,g,a}\}, type, priority)$.

B.4.6 Examples Using Different Attribute Values

Before formally defining the operators, this sub-section presents some examples of how the attributes of a facial signal can be used.

These examples are based on how the EE-FAS works: the current facial signal FS_{cur} is combined with a new facial signals FS_{new} and the attributes of the new facial signal are used to combine both of them.

Selection of a Combination Equation

This example illustrates how an equation is automatically selected to combine two facial signals according to the attributes of one of the facial signals.

$$FS_{new}: \left(\left\{ \begin{array}{l} (C_1, A_{1,1}, T_1, 1, \text{“}multi''\text{, “}add''), \\ (C_1, A_{1,2}, T_2, 1, \text{“}multi''\text{, “}add''), \\ (C_2, A_{2,1}, T_3, 1, \text{“}multi''\text{, “}add'') \end{array} \right\}, \text{“}categorical'', 1 \right)$$
$$FS_{cur}: \left(\left\{ \begin{array}{l} (C_1, A_{1,1}, T_4, 1, \text{“}multi''\text{, “}add''), \\ (C_1, A_{1,3}, T_5, 1, \text{“}multi''\text{, “}add''), \\ (C_3, A_{3,1}, T_6, 1, \text{“}multi''\text{, “}add'') \end{array} \right\}, \text{“}componential'', 6 \right)$$

In this case, the equation C.2 ($FS_1 = FS_r$) is selected due to the attribute *type* = “categorical”. If the *priority* of FS_{new} is equal or superior to 2 the equation C.2 ($FS_1 \cup^X (FS_2 \setminus FS_1) = FS_r$) would be selected because only these two equations would keep the facial component action of FS_{new} unchanged.

$$FS_{new}: \left(\left\{ \begin{array}{l} (C_1, A_{1,1}, T_1, 1, \text{“}multi''\text{, “}add''), \\ (C_1, A_{1,2}, T_2, , 1, \text{“}multi''\text{, “}add''), \\ (C_2, A_{2,1}, T_3, 1, \text{“}multi''\text{, “}add'') \end{array} \right\}, \text{“}componetial'', 3 \right)$$

In this case, with FS_{cur} unchanged, the equation C.3 ($(FS_1 \setminus FS_2) \cup^X (FS_1 \cap FS_2) = FS_r$) is selected due to the attribute *priority* = “3”.

A *categorical* facial signal would be an emotional expression produced by an emotional event, such as an expression of surprise. A *componential* facial signal would be a facial signal corresponding to a communicative function.

Selection of the Operator Indices

This example shows how facial signals can be combined with certain facial components supporting several facial component actions and others supporting only one facial component action to eliminate semantic conflicts.

$$FS_{new}: \left(\left\{ \begin{array}{l} (\text{"}right\_eyebrow\text{"}, \text{"}right\_eyebrow\_frown\text{"}, T_1, \\ 2, \text{"}single\text{"}, \text{"}add\text{"}), \\ (\text{"}mouth\text{"}, \text{"}right\_lip\_corner\_raise\text{"}, T_2, \\ 1, \text{"}multi\text{"}, \text{"}add\text{"}) \end{array} \right\} \\ , \text{"}componential\text{"}, 4 \right)$$

$$FS_{cur}: \left((\left\{ \begin{array}{l} (\text{"}right\_eyebrow\text{"}, \text{"}right\_eyebrow\_raised\text{"}, T_3, \\ 1, \text{"}multi\text{"}, \text{"}add\text{"}), \\ (\text{"}mouth\text{"}, \text{"}viseme\_O\text{"}, T_4, 1, \text{"}multi\text{"}, \text{"}add\text{"}) \end{array} \right\} \\ , \text{"}componential\text{"}, 6 \right)$$

In this case, the equation C.4 is selected:
$(FS_{new} \setminus FS_{cur}) \cup^X (FS_{new} \cap FS_{cur}) \cup^X (FS_{cur} \setminus FS_{new}) = FS_r$.
To build FS_r the sub-sets $(FS_{new} \setminus FS_{cur})$, $(FS_{new} \cap FS_{cur})$, and $(FS_{cur} \setminus FS_{new})$ are built first, then they are unified with the operator $\cup^X$.

To build the three sub-sets, the indices of the operators need to be selected, and this is carried out on an element basis. For the operations $(FS_{new} \setminus FS_{cur})$ and $(FS_{new} \cap FS_{cur})$, the algorithm goes through all the elements of FS_{new}. For the operation $(FS_{cur} \setminus FS_{new})$ the algorithm goes through all the elements of FS_{cur}. For the operation $(FS_{new} \cup FS_{cur})$ the algorithm goes through all the elements of FS_{new} and then through all the elements of FS_{cur}.

For the element $(\text{"}right\_eyebrow\text{"}, \text{"}right\_eyebrow\_frown\text{"}, T_1, 2, \text{"}single\text{"}, \text{"}add\text{"})$ the operators $\setminus_C$, $\cap_C$ and $\cup_C^X$ are used because its attribute TC is equal to "single". For the element $(\text{"}mouth\text{"}, \text{"}right\_lip\_corner\_raise\text{"}, T_2, 1, \text{"}multi\text{"}, \text{"}add\text{"})$ the operators $\setminus_A$, $\cap_A$ and $\cup_A^X$are used because its attribute TC is equal to "multi".

$$(FS_{new} \setminus FS_{cur}) : \left(\begin{array}{l} \left\{ \begin{array}{l} (\text{“}mouth\text{''}, \text{“}right\_lip\_corner\_raise\text{''}, T_2, \\ 1, \text{“}multi\text{''}, \text{“}add\text{''}) \end{array} \right\} \\ , \text{“}componential\text{''}, 4 \end{array} \right)$$

$$(FS_{new} \cap FS_{cur}) : \left(\begin{array}{l} \left\{ \begin{array}{l} (\text{“}right\_eyebrow\text{''}, \text{“}right\_eyebrow\_frown\text{''}, T_1, \\ 2, \text{“}single\text{''}, \text{“}add\text{''}) \end{array} \right\} \\ , \text{“}componential\text{''}, 4 \end{array} \right)$$

$$(FS_{cur} \cap FS_{new}) : \left(\begin{array}{l} \left\{ (\text{“}mouth\text{''}, \text{“}viseme\_O\text{''}, T_4, 1, \text{“}multi\text{''}, \text{“}add\text{''}) \right\} \\ , \text{“}componential\text{''}, 6 \end{array} \right)$$

$$FS_r : \left(\begin{array}{l} \left\{ \begin{array}{l} (\text{“}mouth\text{''}, \text{“}right\_lip\_corner\_raise\text{''}, T_2, \\ 1, \text{“}multi\text{''}, \text{“}add\text{''}) \\ (\text{“}right\_eyebrow\text{''}, \text{“}right\_eyebrow\_frown\text{''}, T_1, \\ 1, \text{“}single\text{''}, \text{“}add\text{''}) \\ (\text{“}mouth\text{''}, \text{“}viseme\_O\text{''}, T_4, 1, \text{“}multi\text{''}, \text{“}add\text{''}) \end{array} \right\} \\ , \text{“}componential\text{''}, 4 \end{array} \right)$$

Blend of the Facial Component Action

This example illustrates the case when a frown is used to communicate an order and another frown is used to emphasise a word.

$$FS_{new} : \left(\begin{array}{l} \left\{ \begin{array}{l} (\text{“}right\_eyebrow\text{''}, \text{“}right\_eyebrow\_frown\text{''}, T_1, \\ 2, \text{“}single\text{''}, \text{“}add\text{''}) \end{array} \right\} \\ , \text{“}componential\text{''}, 4 \end{array} \right)$$

$$FS_{cur} : \left(\begin{array}{l} \left\{ \begin{array}{l} (\text{“}right\_eyebrow\text{''}, \text{“}right\_eyebrow\_frown\text{''}, T_2, \\ 2, \text{“}single\text{''}, \text{“}add\text{''}) \end{array} \right\} \\ , \text{“}componential\text{''}, 6 \end{array} \right)$$

The result of the intersection between FS_{new} and FS_{cur} should be:

$$(FS_{new} \cap FS_{cur}) : \left(\begin{array}{l} \left\{ \begin{array}{l} (\text{“}right\_eyebrow\text{''}, \text{“}right\_eyebrow\_frown\text{''}, T_1, \\ 2, \text{“}single\text{''}, \text{“}add\text{''}) \end{array} \right\} \\ , \text{“}componential\text{''}, 4 \end{array} \right)$$

However, it would be good to have this as a result:

$$(FS_{new} \cap FS_{cur}) : \left(\begin{array}{l} \left\{ \begin{array}{l} (\text{“}right\_eyebrow\text{''}, \text{“}right\_eyebrow\_frown\text{''}, T_{1+2}, \\ 2, \text{“}single\text{''}, \text{“}add\text{''}) \end{array} \right\} \\ , \text{“}componential\text{''}, 4 \end{array} \right)$$

No semantic conflict occurs because it is the same facial component action but both functions of time T_1 and T_2 could be taken into consideration. In this example the intensity of the functions of time will be added together to give the intensity of the facial component action. It is not possible to do this with the current definition of the operator *Intersection* but an improved definition is given later on.

B.4.7 Second Facial Signal Definition

Definition of Facial Signal 2. *The second formal definition of a facial signal is:*

$FS : (\{((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a)_{l,m,i,e,g,a}\}, type, priority)$

where:

- $(C_l, A_{l,m}, X)_{l,m}$ *is unique, X being any function of time. This is to limit conflicting controls of the same animation parameters.*

- $C_l \in FC$. *FC is the set of Facial Components, where each element is unique.*

- $A_{l,m} \in FCA_l$. *FCA_l is the set of the Facial Component Actions affecting one Facial Component C_l, where each element is unique. Each $A_{l,m}$ related to one particular C_l.*

- $C_l = C_{l'}$ *if* $l = l'$.

- $A_{l,m} = A_{l',m'}$ *if* $l = l'$ *and* $m = m'$.

- $T_i = (is, ie, stime, \{(F_n, st_n, et_n)_n\})$. *$T_i$ can be seen as a “function of time” returning the intensity of $A_{l,m}$. is, ie and stime are respectively $A_{l,m}$'s* *intensity at the start, $A_{l,m}$'s intensity at the end and the time at which the $A_{l,m}$ starts. $\{(F_n, st_n, et_n)_n\}$ is a set of functions of time describing a continuous function. st_n is the time at which the function F_n starts and et_n is the time at which the function F_n ends.*

Typically, F_n contains three functions describing the changes of intensity during the attack, the sustain and the decay durations.

- *W_e is the weight associated to $(C_l, A_{l,m}, T_i)$. This weight is used when choices need to be carried out between two facial components or two facial component actions.*
- *TC_g is the type of (facial) component specifying if only one or several facial component actions can occur on this facial component. This attribute can take either the value* single *or* multi *which specify respectively that only one action and several actions can occur on this facial component.*
- *BL_a is the type of blend use between functions of time of a same facial component action. Its values can be:* add, average *or* overwrite.
- type *is the type of facial signal, either* categorical, *specifying that all the component of this facial signal should be kept unchanged, or* componential *when the facial signal can be merge with another facial signal. The value of this attribute, in association with the value of the attribute* priority, *is used to select the equation to combine two facial signals.*
- priority *is a number between 1 to 8 specifying how much influence this facial signal has on the facial signal resulting from the combination. The value of this attribute, in association with the value of the attribute* type, *is used to select the equation to combine two facial signals.*

B.4.8 Second Definition of the Operators *Intersection*

Two Facial Signals:
$FS_1 : (\{((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a)_{l,m,i,e,g,a}\}, type_1, priority_1)$
and $FS_2 : (\{((C_k, A_{k,n}, T_j), W_f, TC_h, BL_b)_{k,n,j,f,h,b}\}, type_2, priority_2)$

Intersection Operator 3.

$FS_1 \cap_C FS_2 = FS_r$

The Intersection$_C$ *operator builds:*
$FS_r : (\{((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z)_{u,v,w,x,y,z}\}, type_3, priority_3)ty_3)$
where $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_C FS_1$ *AND*
$((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_C FS_2$.
$type_3 = type_1$ *and* $priority_3 = priority_1$.

If $m \neq n$

$$\begin{cases} \text{If } W_e \geq W_f \text{ then } v = m,\ w = i,\ x = e,\ y = g \text{ and } z = a \\ \text{If } W_e < W_f \text{ then } v = n,\ w = j,\ x = f,\ y = h \text{ and } z = b \end{cases}$$

If $m = n$:

$$\begin{cases} \text{If } BL_a = overwrite \begin{cases} \text{If } W_e \geq W_f \text{ then } w = i,\ x = e,\ y = g \text{ and } z = a \\ \text{If } W_e < W_f \text{ then } w = j,\ x = f,\ y = h \text{ and } z = b \end{cases} \\ \text{If } BL_a = add \begin{cases} \text{If } W_e \geq W_f \text{ then } T_w = T_i \cup T_j,\ x = e,\ y = g \text{ and } z = a \\ \text{If } W_e < W_f \text{ then } T_w = T_i \cup T_j,\ x = f,\ y = h \text{ and } z = b \end{cases} \end{cases}$$

$T_w = T_i \cup T_j$ *uses the conventional operator union.*

Intersection Operator 4.

$FS_1 \cap_A FS_2 = FS_r$

The $Intersection_A$ *operator builds:*
$FS_r : (\{((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z)_{u,v,w,x,y,z}\}, type_3, priority_3)$
where $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_A FS_1$ *AND*
$((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_A FS_2$.
$type_3 = type_1$ *and* $priority_3 = priority_1$.

$$If\ BL_a = overwrite \begin{cases} If\ W_e \geq W_f\ then\ w = i,\ x = e,\ y = g\ and\ z = a \\ If\ W_e < W_f\ then\ w = j,\ x = f,\ y = h\ and\ z = b \end{cases}$$
$$If\ BL_a = add \begin{cases} If\ W_e \geq W_f\ then\ T_w = T_i \cup T_j,\ x = e,\ y = g\ and\ z = a \\ If\ W_e < W_f\ then\ T_w = T_i \cup T_j,\ x = e,\ y = g\ and\ z = a \end{cases}$$

$T_w = T_i \cup T_j$ *uses the conventional operator union.*

The new definition of all the operators and the set of equations can be found in Appendix C.

B.4.9 Physical Conflicts

To detect and solve physical conflicts, lists of facial component actions, which cannot occur at the same time, are created. Using these lists the system checks if the facial signal resulting from a combination contains any conflict and if a conflict is detected, the *weights* of the conflicting Facial Component Actions are used to select which Facial Component Action should be rejected.

B.5 Animation Discontinuity

In the EE-FAS the formalism is used to combine new facial signals, when new meanings need to be expressed, with the current facial signal displayed on the screen. When conflicts occur, Facial Component Actions can be rejected from the facial signals resulting from the combination. If the rejected facial component actions belong to the facial signal which is displayed, a discontinuity would appear in the animation. This discontinuity is due to the fact that the intensity of the animation parameters controlled by a rejected facial component action would be set to zero independently of their previous intensity values.

The first problem is to detect these discontinuities and the second problem is to resolve them.

B.5.1 Detection of Animation Discontinuities

FS_{new} is the new facial signal, FS_{cur} is the facial signal which is displayed and $FS_{combined}$ is the facial signal resulting of the combination between FS_{new} and FS_{cur}.

With the following operation it is possible to recover facial component actions from FS_{cur}, which have been rejected from $FS_{combined}$ by the combination process: $FS_{rejected} = FS_{cur} \setminus_A FS_{combined}$. The problem is the facial component actions that have been replaced by the same facial component actions but with different functions of time. This type of rejection cannot be detected with the previous operation. One difference between a rejected facial component action and the one which replaced it, is the starting time of the facial component action, *stime*. New operators can be defined to take *stime* into account.

$FS_1 : (\{((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a)_{l,m,i,e,g,a}\}, type_1, priority_1)$
where $T_i = (is_i, ie_i, stime_i, \{(F_n, st_n, et_n)_n\})$ and
$FS_2 : (\{((C_k, A_{k,n}, T_j), W_f, TC_h, BL_b)_{k,n,j,f,h,b}\}, type_2, priority_2)$
where $T_j = (is_j, ie_j, stime_j, \{(F_o, st_o, et_o)_o\})$

In Operator 4. $((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \in_{ST} FS_1$.

$((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \in_{ST} FS_1$ if an element of FS_1 contains the sub-elements $A_{l,m}$ and $stime_i$.

Not In Operator 4. $((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \in_{ST} FS_1$.

$((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \in_{ST} FS_1$ if $A_{l,m}$ and $stime_i$ are not sub-elements of any element of FS_1.

Difference Operator 4. $FS_1 \setminus_{ST} FS_2 = FS_r$.

The $Difference_{ST}$ *operator builds*
$FS_r : (\{((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z)_{u,v,w,x,y,z}\}, type_3, priority_3)$ *where* $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_{ST} FS_1$ *AND* $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \notin_{ST} FS_2$, $type_r = type_1$ *and* $priority_r = priority_1$.

$FS_{rejected} = FS_{displayed} \setminus_{ST} FS_{combined}$ is the set of all the facial component actions which have been rejected from FS_{cur} during the combination process and due to conflicts.

B.5.2 Resolution of Animation Discontinuities

Two types of discontinuity can be found, those due to facial component actions which have been dropped and not replaced, and those due to facial component actions which have been replaced by the same facial component action but with different "functions of time".

Non-replaced Facial Component Action

When a facial component action is dropped from the displayed facial signal its intensity should be decreased to zero in a continuous fashion. A method, shown in Figure B-1, is to re-introduce these facial component actions in the combined facial signal but with a function of time representing the decay of its intensity. Conflicts occur during this period of decay but only during a relatively short length of time, typically few tenth of milliseconds.

Non-replaced Facial Component Action in the EE-FAS

In fact, the EE-FAS implements a different method for the discontinuities due to non-replaced facial component actions. These animation discontinuities are resolved at Level of Control 2, by limiting the speed at which the Abstract Muscle contractions can change. The results presented later in this Chapter are due to

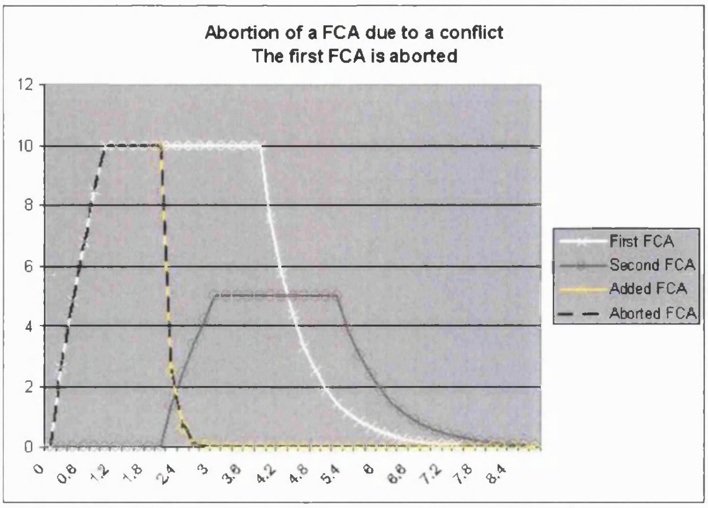

Figure B-1: In this Figure the "First FCA" conflicts with the "Second FCA", therefore the "First FCA" is stopped at the time when the conflict occurs. To eliminate the animation discontinuity, a new FCA, called "Added FCA" is added to the facial signal. The global result is shown by the curve called "Aborted FCA", which is the FCA that has been stop to end the conflict.

this implementation. It should be emphasised that two speed limits have been specified, one for the contraction and one for the relaxation of Abstract Muscles.

Replaced Facial Component Action

A facial component action can be replaced by the same facial component action but with different function of time.In this case, to create a smooth transition between the two controls it is easy to change the starting intensity of the new facial component action. This starting intensity takes the current intensity value of the replaced facial component action, as shown by the Figure B-2.

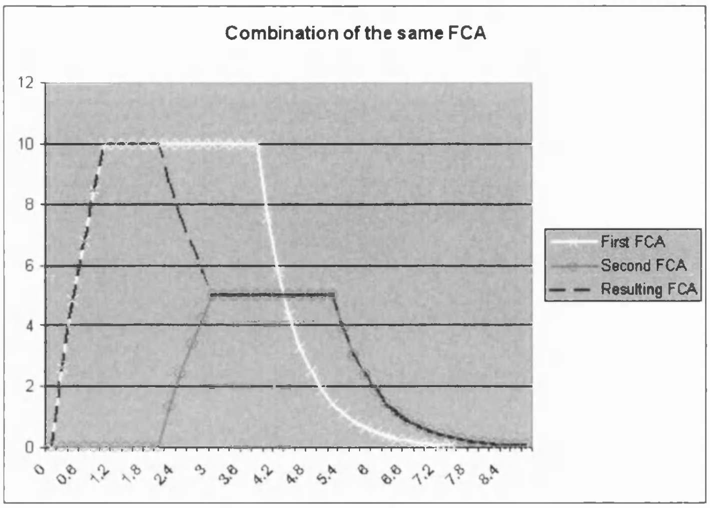

Figure B-2: This Figure shows a blend between two different intensities of the same facial component action. The starting intensity of the "Second FCA" is set to the intensity of the "First FCA" just before its interruption.

B.6 Examples of Facial Signal Combination

B.6.1 Combinations of Different FCAs

Graphs Showing FCA Combinations

Figure B-3 presents five graphs showing different combinations of two facial signals. These signals have been reduced to few Facial Component Actions (FCAs). The first signal, FS_{smile} is composed of "right eyebrow raise" and "right lip corner raise", and the second signal, FS_{angry}, is composed only of "right eyebrow frown". The Facial Component Actions (FCAs) of these signals are interesting because "right eyebrow raise" and "right eyebrow frown" compete to affect the same facial component, which is the "right eyebrow". The facial component action "right lip corner raise" does not compete to affect its facial component but it is influenced by the priority of the signals.

The *priority* and *type* attributes of a facial signal select which equation, defined in Sub-section B.4.4, is used to combine two signals. The *type* "componential" and *priority* "4" selects Equation C.4. With the same *type* and a *priority* "1"

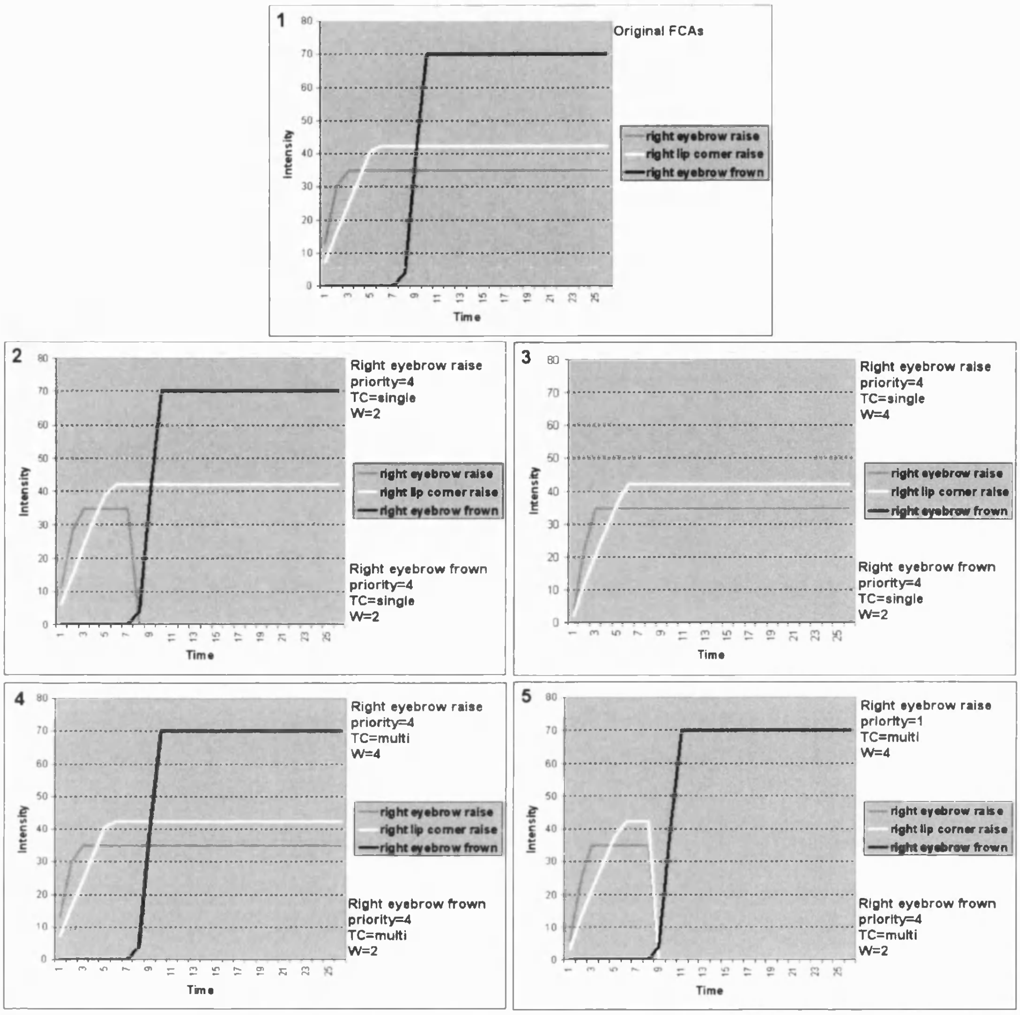

Figure B-3: This figure presents five graphs showing the combinations of two facial signals: "right eyebrow raise" and "right lip corner raise" are part of the same facial signal, and "right eyebrow frown" is part of another signal starting later. Graph 1 shows the Facial Component Actions (FCAs) which are not combined; Graph from 2 to 5 show different set of attributes which change how the FCAs are combined.

Equation C.1 is selected.

Graph 1 on Figure B-3 shows the FCAs not combined. Below are the definitions of the two facial signals.

$$FS_{smile} = \left(\begin{array}{l} \left\{ \begin{array}{l} \left(\begin{array}{l} \text{“right eyebrow”}, \text{“right eyebrow raise”}, 2, \\ \text{“single”}, \text{“overwrite”}, \\ \{(0, 0, 4, \\ \{(f(t)_1, 4, 4.2), (f(t)_2, 4.2, 10), (f(t)_3, 10, 12))\}\} \end{array} \right), \\ \left(\begin{array}{l} \text{“mouth”}, \text{“right lip corner raise”}, 2, \\ \text{“multi”}, \text{“overwrite”}, \\ \{(0, 0, 4, \\ \{(g(t)_1, 4, 4.5), (g(t)_2, 4.5, 10), (g(t)_3, 10, 11))\}\} \end{array} \right) \end{array} \right\}, \\ \text{“componential”}, 4 \end{array} \right)$$

$$FS_{angry} = \left(\begin{array}{l} \left\{ \left(\begin{array}{l} \text{“right eyebrow”}, \text{“right eyebrow frown”}, 2, \\ \text{“single”}, \text{“overwrite”}, \\ \{(0, 0, 6, \\ \{(h(t)_1, 6, 6.2), (h(t)_2, 6.2, 12), (h(t)_3, 12, 14))\}\} \end{array} \right) \right\}, \\ \text{“componential”}, 4 \end{array} \right)$$

Graph 2 shows competition between “right eyebrow raise” and “right eyebrow frown” and the FCA “right eyebrow frown” “wins” over the FAC “right eyebrow raise” due to the position of the operands in the following equation.
$(FS_{angry} \setminus_C FS_{smile}) \cup_C^X (FS_{angry} \cap_C FS_{smile}) \cup_C^X (FS_{smile} \setminus_C FS_{angry}) = FS_r$.

$$FS_{r} = \left(\begin{array}{l} \left\{ \begin{array}{l} \left(\begin{array}{l} \text{“right eyebrow”}, \text{“right eyebrow frown”}, 2, \\ \text{“single”}, \text{“overwrite”}, \\ \{(0, 0, 6, \\ \{(h(t)_1, 6, 6.2), (h(t)_2, 6.2, 12), (h(t)_3, 12, 14))\}\} \end{array} \right), \\ \left(\begin{array}{l} \text{“mouth”}, \text{“right lip corner raise”}, 2, \\ \text{“multi”}, \text{“overwrite”}, \\ \{(0, 0, 4, \\ \{(g(t)_1, 4, 4.5), (g(t)_2, 4.5, 10), (g(t)_3, 10, 11))\}\} \end{array} \right) \end{array} \right\}, \\ \text{“componential”}, 4 \end{array} \right)$$

Graph 3 shows the same situation but the weight of "right eyebrow raise" has been changed from 2 to 4 enabling this FCA to be kept over the "right eyebrow frown".

$(FS_{angry} \setminus_C FS_{smile}) \cup_C^X (FS_{angry} \cap_C FS_{smile}) \cup_C^X (FS_{smile} \setminus_C FS_{angry}) = FS_r.$

$$FS_r = \left(\begin{array}{l} \left\{ \begin{array}{l} \left(\begin{array}{l} \text{“right eyebrow”}, \text{“right eyebrow raise”}, 4, \\ \text{“single”}, \text{“overwrite”}, \\ \{(0, 0, 4, \\ \{(f(t)_1, 4, 4.2), (f(t)_2, 4.2, 10), (f(t)_3, 10, 12))\}\} \end{array} \right), \\ \left(\begin{array}{l} \text{“mouth”}, \text{“right lip corner raise”}, 2, \\ \text{“multi”}, \text{“overwrite”}, \\ \{(0, 0, 4, \\ \{(g(t)_1, 4, 4.5), (g(t)_2, 4.5, 10), (g(t)_3, 10, 11))\}\} \end{array} \right) \end{array} \right\}, \\ \text{“componential”}, 4 \end{array} \right)$$

Graph 4 shows a different combination; in this case the FCAs do not compete for the same facial component because they can both be present at the same time. This type of combination is due to the parameter TC which has the value "multi", enabling a facial component to be affected by multiple facial component actions.

$(FS_{angry} \setminus_A FS_{smile}) \cup_A^X (FS_{angry} \cap_A FS_{smile}) \cup_A^X (FS_{smile} \setminus_A FS_{angry}) = FS_r.$

$$FS_r = \left(\begin{array}{l} \left\{ \begin{array}{l} \left(\begin{array}{l} \text{“right eyebrow”}, \text{“right eyebrow raise”}, 4, \\ \text{“multi”}, \text{“overwrite”}, \\ \{(0, 0, 4, \\ \{(f(t)_1, 4, 4.2), (f(t)_2, 4.2, 10), (f(t)_3, 10, 12))\}\} \end{array} \right), \\ \left(\begin{array}{l} \text{“mouth”}, \text{“right lip corner raise”}, 2, \\ \text{“multi”}, \text{“overwrite”}, \\ \{(0, 0, 4, \\ \{(g(t)_1, 4, 4.5), (g(t)_2, 4.5, 10), (g(t)_3, 10, 11))\}\} \end{array} \right), \\ \left(\begin{array}{l} \text{“right eyebrow”}, \text{“right eyebrow frown”}, 2, \\ \text{“multi”}, \text{“overwrite”}, \\ \{(0, 0, 6, \\ \{(h(t)_1, 6, 6.2), (h(t)_2, 6.2, 12), (h(t)_3, 12, 14))\}\} \end{array} \right) \end{array} \right\}, \\ \text{“componential”}, 4 \end{array} \right)$$

Finally, *Graph 5* shows the combination of facial signals using the equation C.1, which forces the system to display only one signal at one time: $FS_{angry} = FS_r$. When FS_{angry} starts, all the facial component actions of FS_{happy} are stopped. All and the facial component actions of FS_{angry} are displayed and none from FS_{happy}.

$$FS_r = \left(\begin{array}{l} \left\{ \left(\begin{array}{l} \text{“right eyebrow”}, \text{“right eyebrow frown”}, 2, \\ \text{“multi”}, \text{“overwrite”}, \\ \{(0, 0, 6, \\ \{(h(t)_1, 6, 6.2), (h(t)_2, 6.2, 12), (h(t)_3, 12, 14))\}\} \end{array} \right) \right\}, \\ \text{“componential”}, 1 \end{array} \right)$$

Pictures Showing Facial Combinations

Figure B-4 shows six screen-shots of several combinations of the two same facial signals. The two first screen-shots, 1a and 1b, are the two individual facial signals which are combined in the other screen-shots. The facial signals shown in

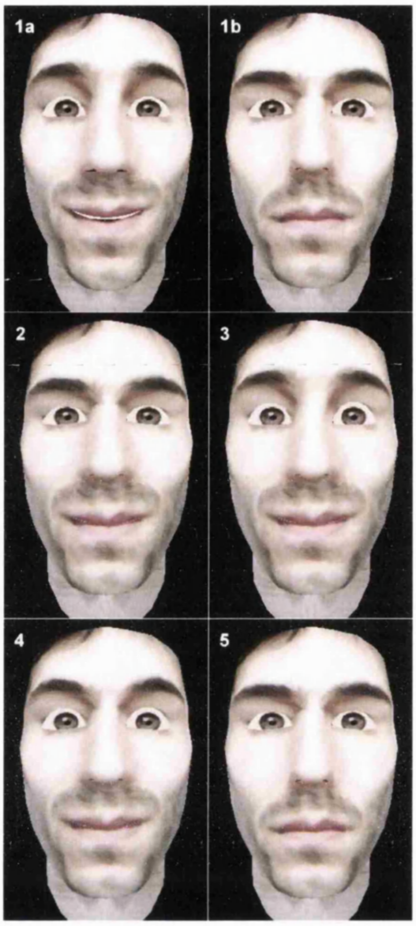

Figure B-4: This figure presents combinations of the two facial signals which are shown in 1a and 1b. According to the attribute values of the facial signals different combinations are carried out: screen-shots 2, 3, 4, 5.

Figure B-4 are screen-shots of the facial component action combinations shown in Figure B-3. The screen-shots show the complete facial signals, which are composed of more facial component actions than the examples in Figure B-3, but the facial component actions in Figure B-3 are all present in the screen-shots. The screen-shots 2 to 5 in Figure B-4 are related respectively to the graphs from 2 to 5 in Figure B-3.

B.6.2 Combination of the Same FCA

Figure B-5 and B-6 shows the combination of two signals FS_{1st} and FS_{2nd} which contain the same facial component actions.

$$FS_{1st} = \left(\begin{array}{l} \left\{ \begin{array}{l} \left(\begin{array}{l} \text{“right eyebrow”}, \text{“right eyebrow raise”}, 2, \\ \text{“multi”}, \text{“overwrite”}, \\ \{(0, 0, 4, \\ \{(f(t)_1, 4, 4.2), (f(t)_2, 4.2, 10), (f(t)_3, 10, 12))\} \end{array} \right), \\ \left(\begin{array}{l} \text{“mouth”}, \text{“right lip corner raise”}, 2, \\ \text{“multi”}, \text{“overwrite”}, \\ \{(0, 0, 4, \\ \{(g(t)_1, 4, 4.5), (g(t)_2, 4.5, 10), (g(t)_3, 10, 11))\} \end{array} \right) \end{array} \right\}, \\ \text{“componential”}, 4 \end{array} \right)$$

$$FS_{2nd} = \left(\begin{array}{l} \left\{ \begin{array}{l} \left(\begin{array}{l} \text{“right eyebrow”}, \text{“right eyebrow raise”}, 2, \\ \text{“multi”}, \text{“overwrite”}, \\ \{(0, 0, 6, \\ \{(h(t)_1, 6, 6.2), (h(t)_2, 6.2, 12), (h(t)_3, 12, 14))\} \end{array} \right), \\ \left(\begin{array}{l} \text{“mouth”}, \text{“right lip corner raise”}, 2, \\ \text{“multi”}, \text{“overwrite”}, \\ \{(0, 0, 6, \\ \{(j(t)_1, 6, 6.5), (j(t)_2, 6.5, 12), (j(t)_3, 12, 13))\} \end{array} \right) \end{array} \right\}, \\ \text{“componential”}, 4 \end{array} \right)$$

$$(FS_{2nd} \setminus_A FS_{1st}) \cup_A^X (FS_{2nd} \cap_A FS_{1st}) \cup_A^X (FS_{1st} \setminus_A FS_{2nd}) = FS_r.$$

The order of the operands in the previous equation is given by the time of arrival of the facial signal.

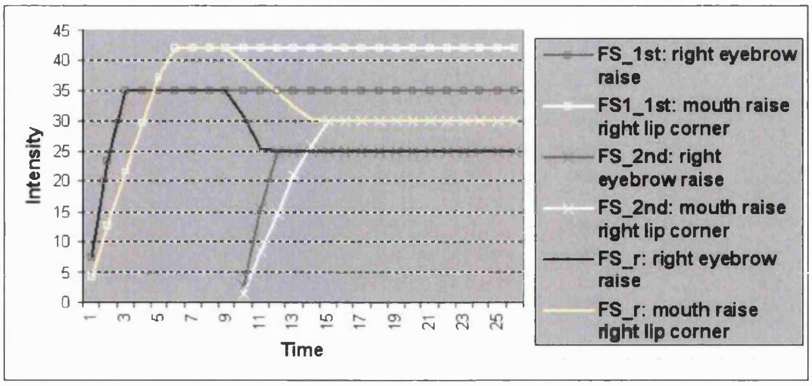

Figure B-5: Blend of different intensities of the same facial component action using the attribute BL = "overwrite".

Type of Blend: Overwrite

Figure B-5 shows the blend of facial component actions using the attribute BL = *overwrite*.

$$FS_r = \left(\begin{array}{l} \left\{ \begin{array}{l} \left(\begin{array}{l} \text{"right eyebrow", "right eyebrow raise"}, 2, \\ \text{"multi", "overwrite"}, \\ \{(f(6), 0, 6, \\ \{(h(t)_1, 6, 6.2), (h(t)_2, 6.2, 12), (h(t)_3, 12, 14))\} \end{array} \right), \\ \left(\begin{array}{l} \text{"mouth", "right lip corner raise"}, 2, \\ \text{"multi", "overwrite"}, \\ \{(g(6), 0, 6, \\ \{(j(t)_1, 6, 6.5), (j(t)_2, 6.5, 12), (j(t)_3, 12, 13))\} \end{array} \right) \end{array} \right\}, \\ \text{"componential"}, 4 \end{array} \right)$$

Type of Blend: Add

Figure B-6 shows the blend of facial component actions using the attribute BL = *add*.

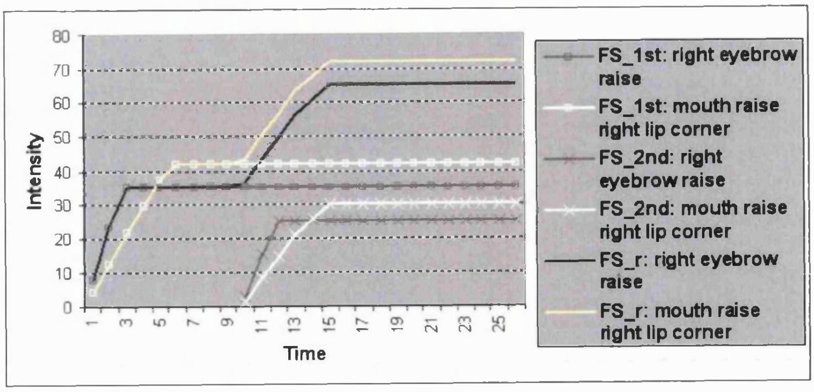

Figure B-6: Blend of different intensities of the same facial component action using the attribute $BL =$ "add".

$$FS_r = \left(\left\{ \begin{array}{l} \left(\begin{array}{l} \text{“right eyebrow”}, \text{“right eyebrow raise”}, 2, \\ \text{“multi”}, \text{“add”}, \\ \{(0, 0, 4, \\ \{((f(t))_1, 4, 4.2), (f(t)_2, 4.2, 10), (f(t)_3, 10, 12)), \\ (0, 0, 6, \\ \{(h(t)_1, 6, 6.2), (h(t)_2, 6.2, 12), (h(t)_3, 12, 14))\} \end{array} \right), \\ \left(\begin{array}{l} \text{“mouth”}, \text{“right lip corner raise”}, 2, \\ \text{“multi”}, \text{“add”}, \\ \{(0, 0, 4, \\ \{(g(t)_1, 4, 4.5), (g(t)_2, 4.5, 10), (g(t)_3, 10, 11)), \\ (0, 0, 6, \\ \{(j(t)_1, 6, 6.5), (j(t)_2, 6.5, 12), ((j(t)_3, 12, 13))\} \end{array} \right) \end{array} \right\}, \\ \text{“componential”}, 4 \right)$$

B.7 Conclusion

This Chapter presented a formalism used to describe facial signals, operators to manipulate facial signals and their components. A set of equations has been selected to represent the combinations between two facial signals. These equations describe a priority mechanism between two facial signals, moving the influence of the resulting facial signal from one to the other. This formalism operates at Level of Control 3.

The description of facial signals is based on how their meanings are carried out, either by the complete facial signal, including all its components, or by some of its components. Such a description enables the system to select the best combination of two facial signals, communicating only one or multiple meanings through the resulting facial signal. The relation between this description and the combination mechanism resolves the semantic conflicts occurring when two facial signals are combined.

This formalism and the combination process have been described in this Chapter supported by concrete examples.

The operators defined in this chapter are used to combine facial signals and to detect animation discontinuities. These operators could also be used in other processes manipulating facial descriptions, such as the detection of facial part free of any activity, which could be used to express new meaning without semantic conflicts.

Appendix C

Definitions and Equations to Combine Facial Signals

This definition provides a set of attributes to the user, describing characteristics of a facial signal and its components, to control the combination process. These attributes related to how the meaning of a facial signal is carried. The set of equation 2, shown later on, describes the different possible combinations between two facial signals.

Definition of Facial Signal 3. *The formal definition of a facial signal is:*

$FS : (\{((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a)_{l,m,i,e,g,a}\}, type, priority)$

where:

- $(C_l, A_{l,m}, X)_{l,m}$ *is unique, X being any function of time. This is to limit conflicting controls of the same animation parameters.*
- $C_l \in FC$. *FC is the set of Facial Components, where each element is unique.*
- $A_{l,m} \in FCA_l$. *FCA_l is the set of the Facial Component Actions affecting one Facial Component C_l, where each element is unique. Each $A_{l,m}$ is related to one particular C_l.*

- $C_l = C_{l'}$ *if* $l = l'$.

- $A_{l,m} = A_{l',m'}$ *if* $l = l'$ *and* $m = m'$.

- $T_i = (is, ie, stime, \{(F_n, st_n, et_n)_n\})$. T_i *can be seen as a "function of time" returning the intensity of* $A_{l,m}$. *is, ie and stime are respectively* $A_{l,m}$*'s intensity at the start,* $A_{l,m}$*'s intensity at the end and the time at which the* $A_{l,m}$ *starts.* $\{(F_n, st_n, et_n)_n\}$ *is a set of functions of time describing a continuous function.* st_n *is the time at which the function* F_n *starts and* et_n *is the time at which the function* F_n *ends.*

  Typically, F_n *contains three functions describing the changes of intensity during the attack, the sustain and the decay durations.*

- W_e *is the weight associated to* $(C_l, A_{l,m}, T_i)$. *This weight is used when choices need to be carried out between two facial components or two facial component actions.*

- TC_g *is the type of (facial) component specifying if only one or several facial component actions can occur on this facial component. This attribute can take either the value* single *or* multi *which specify respectively that only one action and several actions can occur on this facial component.*

- BL_a *is the type of blend use between functions of time of a same facial component action. Its values can be:* add, average *or* overwrite.

- type *is the type of facial signal, either* categorical, *specifying that all the component of this facial signal should be kept unchanged, or* componential *when the facial signal can be merge with another facial signal. The value of this attribute, in association with the value of the attribute* priority, *is used to select the equation to combine two facial signals.*

- priority *is a number between 1 to 8 specifying how much influence this facial signal has on the facial signal resulting from the combination. The value of this attribute, in association with the value of the attribute* type, *is used to select the equation to combine two facial signals.*

Two Facial Signals:
$FS_1 : (\{((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a)_{l,m,i,e,g,a}\}, type_1, priority_1)$
and $FS_2 : (\{((C_k, A_{k,n}, T_j), W_f, TC_h, BL_b)_{k,n,j,f,h,b}\}, type_2, priority_2)$

In Operator 5. $((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \in FS_1$.

$((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \in FS_1$ if FS_1 contains $((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a)$.

In Operator 6. $((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \in_C FS_1$.

$((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \in_C FS_1$ if an element of FS_1 contains the sub-element C_l.

In Operator 7. $((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \in_A FS_1$.

$((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \in_A FS_1$ if an element of FS_1 contains the sub-element $A_{l,m}$.

Not In Operator 5. $((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \notin FS_1$.

$((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \notin FS_1$ if FS_1 does not contain $((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a)$.

Not In Operator 6. $((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \notin_C FS_1$.

$((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \notin_C FS_1$ if C_l is not a sub-element of any element of FS_1.

Not In Operator 7. $((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \notin_A FS_1$.

$((C_l, A_{l,m}, T_i), W_e, TC_g, BL_a) \notin_A FS_1$ if $A_{l,m}$ is not a sub-element of any element of FS_1.

Difference Operator 5. $FS_1 \setminus FS_2$.

The $Difference$ operator builds $\{((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z)_{u,v,w,x,y,z}\}$ where $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in FS_1$ AND $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \notin FS_2$.

This operator is not commutative.

Difference Operator 6. $FS_1 \setminus_C FS_2$.

The $Difference_C$ operator builds $\{((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z)_{u,v,w,x,y,z}\}$ where $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_C FS_1$ AND $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \notin_C FS_2$.

This operator is not commutative.

Difference Operator 7. $FS_1 \setminus_A FS_2$.

The $Difference_A$ *operator builds* $\{((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z)_{u,v,w,x,y,z}\}$ *where* $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_A FS_1$ *AND* $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \notin_A FS_2$.

This operator is not commutative.

XUnion Operator 4. $FS_1 \cup^X FS_2$.

The XUnion *operator builds* $\{((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z)_{u,v,w,x,y,z}\}$ *where* $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in FS_1$ *XOR* $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in FS_2$.

XUnion Operator 5. $FS_1 \cup^X_C FS_2$.

The XUnion *operator builds* $\{((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z)_{u,v,w,x,y,z}\}$ *where* $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_C FS_1$ *XOR* $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_C FS_2$.

This operator is commutative.

XUnion Operator 6. $FS_1 \cup^X_A FS_2$.

The XUnion *operator builds* $\{((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z)_{u,v,w,x,y,z}\}$ *where* $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_A FS_1$ *XOR* $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_A FS_2$.

This operator is commutative.

Intersection Operator 5.

$FS_1 \cap_C FS_2 = FS_r$

The $Intersection_C$ *operator builds:*
$FS_r : (\{((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z)_{u,v,w,x,y,z}\}, type_3, priority_3)$
where $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_C FS_1$ *AND*
$((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_C FS_2$.
$type_3 = type_1$ *and* $priority_3 = priority_1$.

If $m \neq n$

$$\begin{cases} \textit{If } W_e \geq W_f \textit{ then } v = m,\ w = i,\ x = e,\ y = g \textit{ and } z = a \\ \textit{If } W_e < W_f \textit{ then } v = n,\ w = j,\ x = f,\ y = h \textit{ and } z = b \end{cases}$$

If $m = n$:

$$\begin{cases} \textit{If } BL_a = \textit{overwrite} \begin{cases} \textit{If } W_e \geq W_f \textit{ then } w = i,\ x = e,\ y = g \textit{ and } z = a \\ \textit{If } W_e < W_f \textit{ then } w = j,\ x = f,\ y = h \textit{ and } z = b \end{cases} \\ \textit{If } BL_a = \textit{add} \begin{cases} \textit{If } W_e \geq W_f \textit{ then } T_w = T_i \cup T_j,\ x = e,\ y = g \textit{ and } z = a \\ \textit{If } W_e < W_f \textit{ then } T_w = T_i \cup T_j,\ x = f,\ y = h \textit{ and } z = b \end{cases} \end{cases}$$

$T_w = T_i \cup T_j$ *uses the conventional operator union.*

Intersection Operator 6.

$FS_1 \cap_A FS_2 = FS_r$

The $Intersection_A$ operator builds:
$FS_r : (\{((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z)_{u,v,w,x,y,z}\}, type_3, priority_3)$
where $((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_A FS_1$ *AND*
$((C_u, A_{u,v}, T_w), W_x, TC_y, BL_z) \in_A FS_2$.
$type_3 = type_1$ *and* $priority_3 = priority_1$.

$$\text{If } BL_a = overwrite \begin{cases} \text{If } W_e \geq W_f \text{ then } w = i,\ x = e,\ y = g \text{ and } z = a \\ \text{If } W_e < W_f \text{ then } w = j,\ x = f,\ y = h \text{ and } z = b \end{cases}$$
$$\text{If } BL_a = add \begin{cases} \text{If } W_e \geq W_f \text{ then } T_w = T_i \cup T_j,\ x = e,\ y = g \text{ and } z = a \\ \text{If } W_e < W_f \text{ then } T_w = T_i \cup T_j,\ x = e,\ y = g \text{ and } z = a \end{cases}$$

$T_w = T_i \cup T_j$ *uses the conventional operator union.*

Equations for Combining Facial Displays 2.

$$FS_1 = FS_r \tag{C.1}$$
$$FS_1 \cup^X (FS_2 \setminus FS_1) = FS_r \tag{C.2}$$
$$(FS_1 \setminus FS_2) \cup^X (FS_1 \cap FS_2) = FS_r \tag{C.3}$$
$$(FS_1 \setminus FS_2) \cup^X (FS_1 \cap FS_2) \cup^X (FS_2 \setminus FS_1) = FS_r \tag{C.4}$$
$$(FS_2 \setminus FS_1) \cup^X (FS_2 \cap FS_1) \cup^X (FS_1 \setminus FS_2) = FS_r \tag{C.5}$$
$$(FS_2 \setminus FS_1) \cup^X (FS_2 \cap FS_1) = FS_r \tag{C.6}$$
$$FS_2 \cup^X (FS_1 \setminus FS_2) = FS_r \tag{C.7}$$
$$FS_2 = FS_r \tag{C.8}$$

where $\cup^X$ *could be either* $\cup^X_C$ *or* $\cup^X_A$, $\setminus$ *could be either* $\setminus_C$ *or* $\setminus_A$ *and* $\cap$ *could be either* $\cap_C$ *or* $\cap_A$. *In fact the type of operator used is not specified at a facial signal level but on a facial component action basis.*

Equation C.1 describes a case where the FS_1 has a complete influence on the result, which means that FS_1 has total priority over FS_2. As a logical regression of FS_1's priority, Equation C.2 enables FS_2 to contribute to the resulting facial signal by adding its elements which are not in FS_1 and all the elements of FS_1 are conserved unchanged. Equation C.3 provides even more influence to FS_2, enabling it to play a role by eventually modifing the elements in common with FS_1. Equation C.4 enables FS_2 to play a role based on the common elements with FS_1 but also to add its elements which are not in FS_1 to the resulting facial signal. Equations C.4 and C.5 are different due to the non-commutativity of the operator *Intersection.* Due to this characteristic Equation C.5 provides more influence to FS_2. The rest of the equations can be described in the same way but inverting the priorities between FS_1 and FS_2.

Appendix D

Publications

- A Layered Dynammic Emotion Representaiton for the Creation of Complex Facial Expressions
  Mr. Emmanuel Tanguy, Prof. Philip Willis and Dr. Joanna J. Bryson.
  Department of Computer Science; University of Bath; Bath BA2 7AY, England.
  4th International Workshop, IVA 2003;Kloster Irsee, Germany, 15-17 of September 2003. Proceedings pages: 101-105
- The Role of Emotions in Modular Intelligent Control
  Joanna J Bryson, Emmanuel Tanguy and Philip Willis.
  Department of Computer Science; University of Bath; Bath BA2 7AY, England.
  AISB Quarterly, Number 117, page 1 and page 6.
- Building Expression into Virtual Characters
  V. Vinayagamoorthy[1], M. Gillies[1], A. Steed[1], E. Tanguy[2], X. Pan[1], C. Loscos[1], and M. Slater[1]
  [1]Department of Computer Science, University College London, London, UK
  [2]Department of Compter Science, University of Bath, Bath, UK
  Eurographics 2006, STAR – State of The Art Report, Vienna, to be published.

Bibliography

Albrecht, I. (2005, December). *-Faces and Hands- Modeling and animating anatomical and photorealistic models with regard to the communicative competence of virtual humans.* Ph. D. thesis, Universität des Saarlandes.

Albrecht, I., J. Haber, K. Kähler, M. Shröder, and H.-P. Seidel (2002). "May I talk to you? :-)" — Facial Animation from Text. In *10th Pacific Conference on Computer Graphics and Applications*, pp. 77. IEEE Computer Society.

Albrecht, I., M. Schröder, J. Haber, and H.-P. Seidel (2005). Mixed feelings: Expression of non-basic emotions in a muscle-based talking head. *Journal of Virtual Reality 8*, 201–212. Special issue on Language, Speech & Gesture.

André, E., M. Klesen, P. Gebhard, S. Allen, and T. Rist (1999, October). Integrating models of personality and emotions into lifelike character. In *Proceeding of the Workshop on Affect in Interaction — Towards a new Generation of Interfaces.*, Siena, Italy, pp. 136–149.

Arnold, M. B. (1960). *Emotion and Personality*, Volume 1. New York: Columbia University Press.

Badler, N., J. Allbeck, L. Zhao, and M. Byun (2002). Representing and parameterizing agent behaviors. In *Proceeding Computer Animation*, Geneva, Switzerland, pp. 133–143. IEEE Computer Society June 2002.

Bartneck, C. (2002). Integrating the OCC model of emotions in embodied characters. In *Proceedings of the Workshop on Virtual Conversational Characters: Applications, Methods, and Research Challenges*, Melbourne.

Bavelas, J. B. and N. Chovil (1997). Faces in dialogue. See Russel and Fernández-Dols (1997), pp. 334–346.

Bavelas, J. B. and N. Chovil (2000). Visible acts of meaning: An integrated message model of language in face-to-face dialogue. *Journal of Language and Social Psychology 19*(2), 163–194.

Bui, T. D., D. Heylen, and A. Nijholt (2004). Combination of facial movements on a 3d talking head. In *Proceedings Computer Graphics International 2004*, pp. 284–291. IEEE Computer Society. Crete, Greece June 2004.

Cañamero, D. (1998). Issues in the design of emotional agents. In *Emotional and Intelligent: The Tangled Knot of Cognition*, Papers from the 1998 AAAI Fall Symposium. Menlo Park, pp. 49–54. AAAI Press.

Cañamero, D. (2003). Designing emotions for activity selection in autonomous agents. In R. Trappl, P. Petta, and S. Payr (Eds.), *Emotions in Humans and Artifacts*, pp. 115–148. The MIT Press.

Carolis, B. D., V. Carofiglio, and C. Pelachaud (2002, July). From discourse plans to believable behavior generation. In *International Natural Language Generation*, New York, USA, pp. 65-72.

Cassell, J., H. H. Vilhjálmsson, and T. Bickmore (2001). BEAT: the Behavior Expression Animation Toolkit. In *ACM SIGGRAPH,Computer Graphics Proceedings*, pp. 477–486.

Chovil, N. (1992). Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction 25*, 163–194.

Chovil, N. (1997). Facing others: A social communicative perspective on facial displays. See Russel and Fernández-Dols (1997), pp. 321–333.

Cohen, M. M. and D. W. Massaro (1994). Modeling coarticulation in synthetic visual speech. In N. Magnenat Thalmann and D. Thalmann (Eds.), *Models and Techniques in Computer Animation*, pp. 139–156. Tokyo: Springer.

Collier, G. (1885). *Emotional expression.* Lawrence Erlbaum Associates.

CSTR (2006). The festival speech synthesis system. Online. The Centre for Speech Technology Research: http://www.cstr.ed.ac.uk/projects/festival/.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain.* Putnam.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness.* Heinemann.

Darwin, C. (1979). *Expression of the emotions in man and animals.* J Friedmann.

De Rosis, F., C. Pelachaud, I. Poggi, V. Carofiglio, and B. D. Carolis (2003). From Greta's mind to her face: Modelling the dynamics of affective states in a conversational embodied agent. *International Journal Human-Computer Studies: Special issue on "Application of Affective Computing in HCI" 59*(1-2), 81–118.

DeCarlos, D., C. Revilla, and M. Stone (2002). Making discourse visible: Coding and animating conversational facial displays. In *Computer Animation 2002*, pp. 11–16. IEEE Computer Society.

Delgado-Mata, C. (2004). *Emotion signalling in multiple intelligent virtual agents for believable artificial animals.* Ph. D. thesis, University of Salford, uk.

Delgado-Mata, C. and R. S. Aylett (2004). Emotion and action selection: regulating the collective behaviour of agents in virtual environments. In *AAMAS ' 04: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, Washington, DC, USA, pp. 1304–1305. IEEE Computer Society.

Descartes, R. (1973/1911/1649). Passions of the soul. In E. S. Haldane and G. R. T. Ross (Eds.), *The philosophical work of descartes*, pp. 331–427. Cambridge university press.

Edwards, D. C. (1999). *Motivation & Emotion.* SAGE publications.

Egges, A., S. Kshirsagar, and N. Magnenat-Thalmann (2004, January). Generic personality and emotion simulation for conversational agents. *Computer Animation and Virtual Worlds 15*, 1–13.

Egges, A., S. Kshirsagar, X. Zhang, and N. Magnenat-Thalmann (2003). Emotional communication with virtual humans. In *International Conference on Multi-Media Modeling*, pp. 243–263.

Ekman, P. (1979). *About brows: Emotional and conversational signals*, Chapter 3.1, pp. 169–202. Cambridge University Press.

Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science 3*, 34–38.

Ekman, P. (1999a). Basic Emotions. In T. Dalgleish and M. Power (Eds.), *Handbook of Cognition and Emotion*, pp. 45–60. John Wiley & Sons Ltd.

Ekman, P. (1999b). Facial Expressions. In T. Dalgleish and M. Power (Eds.), *Handbook of Cognition and Emotion*, pp. 301–302. John Wiley & Sons Ltd.

Ekman, P. and R. J. Davidson (1994). *The nature of emotion: Fundamental questions.* Oxford University Press.

Ekman, P. and W. V. Friesen (1969). *The repertoire of non-verbal behaviour categories, origins usage and coding.*, pp. 49–98. Semiotica 1, Pergamon Press.

Ekman, P. and W. V. Friesen (1978). *Facial Action Coding System* (Consulting Psychologist Press ed.).

Eysenck, M. W. and M. T. Keane (2002). *Cognitive Psychology: A student's handbook* (Fourth ed.). Psychology Press Ltd.

Fehr, B. and J. A. Russell (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology. General 113*, 464–486.

Fernández-Dols, J. M. and M.-A. Ruiz-Belda (1997). Spontaneous facial behavior during intense emotional episodes: Artistic truth and optional truth. See Russel and Fernández-Dols (1997), pp. 255–274.

Fox, E., V. Lester, R. Russo, R. J. B. andA. Pichler, and K. Dutton (2000). Facial expression of emotion: Are angry faces detected more efficiently? *Cognition and Emotion 14*(1), 61–92.

Fridlund, A. J. (1997). The new ethology of human facial expressions. See Russel and Fernández-Dols (1997), pp. 103–129.

Frijda, N. H. (1986). *The Emotions.* Cambridge University Press.

Frijda, N. H. (2000). The Psychologists' Point of View. In M. Lewis and J. M. Haviland-Jones (Eds.), *Handbook of Emotions*, pp. 59–74. The Guilford Press.

Frijda, N. H. and A. Tcherkassof (1997). Facial expression as modes of action readiness. See Russel and Fernández-Dols (1997), pp. 78–102.

Fuchs, T., J. Haber, and H.-P. Seidel (2004). MIMIC — a language for specifying facial animations. In *Proceedings of WSCG 2004*, pp. 71–78.

Ginsburg, G. P. (1997). Faces: An epilogue and reconceptualization. See Russel and Fernández-Dols (1997), pp. 349–382.

Gratch, J. (2000). Émile: Marshalling passions in training and education. In C. Sierra, M. Gini, and J. S. Rosenschein (Eds.), *Proceedings of the Fourth International Conference on Autonomous Agents*, Barcelona, Catalonia, Spain, pp. 325–332. ACM Press.

Gratch, J. and S. Marsella (2004a, December). A domain-independent framework for modeling emotion. *Cognitive Systems Research 5*(4), 269–306.

Gratch, J. and S. Marsella (2004b). Evaluating the modeling and use of emotion in virtual humans. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, Volume 1, pp. 320–327. IEEE Computer Society.

Gratch, J., J. Rickel, E. Andre, N. Badler, J. Cassell, and E. Petajan (2002, July–August). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems 17*(4), 54–63.

Gustavsson, C., L. Strindlund, E. Wiknertz, S. Beard, Q. Huynh, A. Marriott, and J. Stallo (2001, 10). VHML. Working Draft v0.3 October 21st 2001.

Hartung, F., P. Eisert, and B. Girod (1998). Digital watermarking of MPEG-4 facial animation parameters. *IEEE Computers and Graphics 22*(4), 425–435.

Izard, C. E. (1971). *The face of emotion.* Appleton-Century-Crofts.

Izard, C. E. (1977). *Human emotions.* Plenum Press.

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review 100*(1), 68–90.

Izard, C. E. (1997). Emotions and facial expression: A perspective from differential emotions theory. See Russel and Fernández-Dols (1997), pp. 57–77.

Kalra, P., A. Mangili, N. Magnenat-Thalmann, and D. Thalmann (1991). SMILE: A multilayered facial animation system. In K. T. L (Ed.), *Proceedings IFIP conference on Modelling in Computer Graphics*, pp. 189–198. Springer.

Kshirsagar, S. and N. Magnenat-Thalmann (2002, June). A multilayer personality model. In *Proceedings of* 2^{nd} *International Symposium on Smart Graphics*, Hawthorne, New York, pp. 107–115. ACM Press.

Kurlander, D., T. Skelly, and D. Salesin (1996). Comic chat. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, New York, NY, USA, pp. 225–236. ACM Press.

Latta, C., N. Alvarado, S. S. Adams, and S. Burbeck (2002). An expressive system for endowing robots or animated characters with affective facial displays. In *The 2002 conference of the British Society for Artificial Intelligence and the Simulation of Behavior.*

Lazarus, R. S. (1991). *Emotion and adaptation.* Oxford University Press.

LeDoux, J. E. (1999). *The emotional brain : the mysterious underpinnings of emotional life.* Phoenix.

Oatley, K. and J. M. Jenkins (1996). *Understanding Emotions.* Blackwell Publishers Ltd.

Ortony, A. (2001). *On making believable emotional agents believable.* Cambridge: MIT Press.

Ortony, A., G. L. Clore, and A. Collins (1990/1988). *The cognitive structure of emotions.* Cambridge University Press.

Ortony, A. and T. J. Turner (1990). What's Basic About Basic Emotions? *Psychological Review 97*(3), 315–331.

Paiva, A., J. Dias, D. Sobral, R. Aylett, P. Sobreperez, S. Woods, C. Zoll, and L. Hall (2004). Caring for agents and agents that care: Building empathic relations with synthetic agents. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, Washington, DC, USA, pp. 194–201. IEEE Computer Society.

Paradiso, A. (2002). An algebra for combining MPEG-4 compliant facial animations. In H. Prendinger (Ed.), *Online Proceedings; International Workshop on Lifelike Animated Agents: Tools, Affective Functions, and Applications.* Tokyo, Japan, 19th of August 2002.

Parke, F. I. and K. Waters (1996). *Computer Facial Animation.* A K Peters Ltd.

Parkinson, B. (1995). *Ideas and realities of emotion.* Routledge.

Pelachaud, C., N. Badler, and M. Steedman (1991). Linguistic issues in facial animation. In N. Magnenat-Thalmann and D. Thalmann (Eds.), *Computer Animation'91*, pp. 15–30. Springer-Verlag.

Pelachaud, C. and M. Bilvi (2003). Computational model of believable conversational agents. In M.-P. Huget (Ed.), *Communication in Multiagent Systems*, pp. 300–317. Springer.

Pelachaud, C. and I. Poggi (1998, October). Facial performative in a conversational system. In *Proceedings of the WECC98.* Workshop on Embodied Conversational Characters.

Picard, R. W. (1997). *Affective Computing.* The MIT Press, Cambridge, Massachusetts.

Pirker, H. and B. Krenn (2002). Assessment of markup languages for avatars, multimedia and multimodal system. Technical report, ÖFAI.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik and H. Kellerman (Eds.), *Emotion: Theory, research, and experience*, pp. 3–33. Academic press, inc.

Poggi, I. and C. Pelachaud (2000). Emotional meaning and expression in animated faces. In *Affective interactions: towards a new generation of computer interfaces*, pp. 182–195. Springer-Verlag New York, Inc.

Raouzaiou, A., K. Karpouzis, and S. D. Kollias (2003). Online gaming and emotion representation. In N. N. García, J. M. Martínez, and L. Salgado (Eds.), *VLBV*, Volume 2849 of *Lecture Notes in Computer Science*, pp. 298–305. Springer.

Reilly, W. S. and J. Bates (1992). Building emotional agents. Technical Report CMU-CS-92-143, Pittsburg, PA, USA.

Reilly, W. S. N. (1996). *Believable social and emotional agents.* Ph. D. thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh.

Roseman, I. J., A. Antoniou, and P. E. Jose (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition and emotion 10*(3), 241–277.

Russel, J. A. and J. M. Fernández-Dols (1997). *The Psychology of Facial Expression.* Cambridge University Press.

Russell, J. A. (1997). Reading emotion from and into faces: Resurrecting a dimensional-contextual perspective. See Russel and Fernández-Dols (1997), pp. 295–320.

Scherer, K. R. (1988). *Facets of emotion: Recent research.* Lawrence Erlbaum Associates.

Sendak, M. (1963). *Where the wild things are.* Harper Collins Publisher.

Sizer, L. (2000, December). Towards a computational theory of mood. *The British Journal for the Philosophy of Science 51*(4), 743–769.

Sloman, A. (2000). Architectural requirements for human-like agents both natural and artificial. (what sorts of machines can love?). In K. Dautenhahn (Ed.), *Human Cognition And Social Agent Technology*, pp. 163–195. John Benjamins Publishing Company.

Sloman, A. (2001a). Beyond shallow models of emotions. *Cognitive Processing 2*(1), 177–198.

Sloman, A. (2001b, March). Varieties of affect and the CogAff architecture schema. In *Proceedings Symposium on Emotion, Cognition, and Affective Computing AISB'01 Convention*, York, pp. 39–48.

Sloman, A. (2003). How many separately evolved emotional beasties live within us? In *Emotions in Humans and Artifacts*, pp. 35–114. The MIT Press.

Smith, C. A. and H. S. Scott (1997). A componential approach to the meaning of facial expressions. See Russel and Fernández-Dols (1997), pp. 295–320.

Tanguy, E. A. R., P. J. Willis, and J. J. Bryson (2003, September). A Layered Dynamic Emotion Representation for the Creation of Complex Facial Animation. In T. Rist, R. Aylett, D. Ballin, and J. Rickel (Eds.), *Intelligent Virtual Agents*, pp. 101–105. Springer.

Thayer, R. E. (1996). *The origin of everyday moods.* Oxford University Press.

The Duy Bui (2004). *Creating emotions and facial expressions for embodied agents.* Ph. D. thesis, University of Twente.

Thórisson, K. R. (1999). A mind model for multimodal communicative creatures & humanoids. *International Journal of Applied Artificial Intelligence 13*(4/5), 519–538.

Thórisson, K. R. (2004). Communicative machines: Psyclone. http://www.cmlabs.com/psyclone.

Thórisson, K. R., C. Pennock, T. List, and J. DiPirro (2004). Artificial intelligence in computer graphics: A constructionist approach. *Computer Graphics Quarterly*, 26–30. New York: Association for Computing Machinery.

Tomkins, S. S. (1980). Affect as amplification: some modification in theory. In R. Plutchik and H. Kellerman (Eds.), *Emotion: Theory, research, and experience*, pp. 3–33. Academic Press, Inc.

Velásquez, J. D. (1997). Modeling emotions and other motivations in synthetic agents. In *proceedings of Fourteenth AAAI National Conference Artificial Intelligence*, pp. 10–15.

Wehrle, T., S. Kaiser, S. Schmidt, and K. R. Scherer (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology 78*(1), 105–119.